

# Knowledge and Security

Riccardo Pucella\*

## Abstract

Epistemic concepts, and in some cases epistemic logic, have been used in security research to formalize security properties of systems. This survey illustrates some of these uses by focusing on confidentiality in the context of cryptographic protocols, and in the context of multi-level security systems.

*Security: the state of being free from danger or threat.*

*(New Oxford American Dictionary)*

## 1 Introduction

A persistent intuition in some quarters of the security research community says that epistemic logic and, more generally, epistemic concepts are useful for reasoning about the security of systems. What grounds this intuition is that much work in the field is based on epistemic concepts—sometimes explicitly, but more often implicitly, by and large reinventing possible-worlds semantics for knowledge and belief.<sup>1</sup>

Reasoning about the security of systems in practice amounts to establishing that those systems satisfy various security properties. A security property, roughly speaking, is a property of a system stating that the system is not vulnerable to a particular threat. Threats, in this context, are generally taken to be attacks by agents intent on subverting the system.

While what might be considered a threat—and therefore what security properties are meant to protect against—is in the eye of the beholder, several properties have historically been treated as security properties:

- **Data Confidentiality:** only authorized agents should have access to a piece of data; more generally, only authorized agents should be able to infer any information about a piece of data.
- **Data Integrity:** only authorized agents should have access to alter a piece of data.

---

\* Author's email: [riccardo@acm.org](mailto:riccardo@acm.org). To appear in *Handbook of Logics for Knowledge and Belief*.

<sup>1</sup>This chapter assumes from the reader a basic knowledge of epistemic logic and its Hintikka-style possible-worlds semantics; see §6 for references. Furthermore, to simplify the exposition, the term *epistemic* is used to refer both to knowledge and to belief throughout.

- **Agent Authentication:** an agent should be able to prove her identity to another agent.
- **Data Authentication:** an agent should be able to determine the source of a piece of data.
- **Anonymity:** the identity of an agent or the source of a piece of data should be kept hidden except from authorized agents.
- **Message Non-repudiation:** the sender of a message should not be able to deny having sent the message.

These properties may seem intuitive on a first reading but they are vague and depend on terms that require clarification: *secret, authorized, access to a piece of data, source, identity*.

Epistemic concepts come into play when defining many of the terms that appear in the statements of security properties. Indeed, those terms can often be usefully understood in terms of knowledge: confidentiality can be read as *no agent except for authorized agents can know a piece of information*; authentication as *an agent knows the identity of the agent with whom she is interacting, or an agent knows the identity of the agent who sent the information*; anonymity as *no one knows the identity of the agent who performed a particular action*; and so on. While it is not the case that every security property can be read as an epistemic property, enough of them can to justify studying them *as* epistemic properties.

Epistemic logic and epistemic concepts play two roles in security research:

- **Definitional:** they are used to formalize security properties and concepts, and provide a clear semantic grounding for them. Epistemic logic may be explicitly used as an explanatory and definitional language for properties of interest.
- **Practical:** they are used to derive verification and enforcement techniques for security properties, that is, to either establish that a security property is true in a system, or to force a security property to be true in a system.

It is fair to say that after nearly three decades of research, epistemic logic has had several successes on the definitional front and somewhat fewer on the practical front. This is perhaps not surprising. While epistemic logic and epistemic concepts are well suited for definitions and for describing semantic models, verification of epistemic properties tends to be expensive, and tools for the verification of security properties in practice often approximate epistemic properties using properties that are easier to check, such as safety properties.<sup>2</sup>

This chapter illustrates the use of epistemic logic and epistemic concepts for reasoning about security through the study of a specific security property, confidentiality. Not only is confidentiality a prime example of the use of knowledge to make a security property precise, but it has also been studied extensively from several perspectives. Moreover, many

---

<sup>2</sup>A safety property is a property of the form *a bad state is never reached in any execution of the system*, for some definition of *bad state*. A safety property can be checked by examining every possible execution independently of any other; in contrast, checking an epistemic property requires examining every possible execution in the context of all other possible executions.

of the issues arising while studying confidentiality also arise for other security properties with an epistemic flavor.

Confidentiality is explored in two contexts: cryptographic protocols in §2, and multi-level security systems in §3. Cryptographic protocols are communication protocols that use cryptography to protect information exchanged between agents in a system. While it may seem simple enough for Alice to send a confidential message to Betty by encrypting it, Alice and Betty need to share a common key for this to work. How is such a key distributed before communication may take place? Most cryptographic protocols involve key creation and distribution, and these are notoriously difficult to get right. Key distribution also forces the consideration of authentication as an additional security property. Cryptographic protocol analysis is the one field of security research that has explicitly and extensively used epistemic logic, and the bulk of this chapter is dedicated to that topic.

In multi-level security systems, one is generally interested in confidentiality guarantees even when information is used or released within the system during a computation. The standard example is that of a centralized system where agents have different security clearances and interact with data with different security classifications; the desired confidentiality guarantees ensure that classified data, no matter how it is manipulated by agents with an appropriate security clearance, never flows to an agent that does not have an appropriate security clearance. Most of the work in this field of security research uses epistemic concepts implicitly—the models use possible-worlds definitions of knowledge, but no epistemic logic is introduced. All reasoning is semantic reasoning in the models.

Security properties other than confidentiality are briefly discussed in §4. The chapter concludes in §5 with some personal views on the use of epistemic logic and epistemic concepts in security research. My observations should not be particularly controversial, but my main conclusion remains that progress beyond the current state of the art in security research—at least in security research that benefits from epistemic logic—will require a deeper understanding of resource-bounded knowledge, which is itself an active research area in epistemic logic.

All bibliographic references are postponed to §6, where full references and additional details are given for topics covered in the main body of the chapter. It is worth noting that the literature on reasoning about security draws from several fields besides logic. For instance, much of the research on cryptographic protocols derives from earlier work in distributed computing. Similarly, recent research both on cryptographic protocol analysis and on confidentiality in multi-level systems is based on work in programming language semantics and static analysis. The interested reader is invited to follow the references given in §6 for details.

## 2 Cryptographic Protocols

Cryptographic protocols are communication protocols—rules for exchanging messages between agents—that use cryptography to achieve a security goal such as authenticating one agent to another, or exchanging confidential messages.

Cryptographic protocols are a popular object of study for several reasons. First, they are concrete—they correspond to actual artifacts implemented and used in practice. Sec-

ond, their theory extends that of distributed protocols and network protocols in general, which are themselves thoroughly studied.

Cryptographic protocols have characteristics that distinguish them from more general communication protocols. In particular, they

- (1) enforce security properties;
- (2) rely on cryptography;
- (3) execute in the presence of attackers that might attempt to subvert them.

Protocols can be analyzed concretely or symbolically. The concrete perspective views protocols as exchanging messages consisting of sequences of bits and subject to formatting requirements, which is the perspective used in most network protocols research. The symbolic perspective views protocols as exchanging messages consisting of symbols in some formal language, which is the perspective used in most distributed protocols research. The focus of this section is on symbolic cryptographic protocols analysis.

## 2.1 Protocols

A common notation for protocols is to list the sequence of messages exchanged between the parties involved in the protocol, since the kinds of protocols studied rarely involve complex control flow.

A simple protocol between Alice and Betty (represented by  $A$  and  $B$ ) in which Alice sends message  $m_1$  to Betty and Betty responds by sending message  $m_2$  to Alice would be described by:

$$\begin{array}{lll} 1. & A \longrightarrow B & : \quad m_1 \\ 2. & B \longrightarrow A & : \quad m_2 \end{array} \quad (1)$$

The message sequence notation takes a global view of the protocol, describing the protocol from the outside, so to speak. An alternate way to describe a protocol is to specify the roles of the parties involved in the protocol. For protocol (1), for instance, there are two roles: the initiator role, who sends message  $m_1$  to the receiver and waits for a response message, and the receiver role, who waits for a message to arrive from the initiator and responds with  $m_2$ .

A protocol executes in an environment, which details anything relevant to the execution of said protocol, such as the agents participating in the protocol, whether other instances of the protocol are also executed concurrently, the possible attackers and their capabilities. The result of executing a protocol in a given environment can be modeled by a set of traces, where a trace corresponds to a possible execution of the protocol. A trace is a sequence of global states. A global state records the local state of every agent involved in the protocol, as well as the state of the environment. This general description is compatible with most representations in the literature, and can be viewed as a Kripke structure by defining a suitable accessibility relation over the states of the system.

To illustrate protocols in general, and initiate the study of confidentiality, here are two simple protocols that achieve a specific form of confidentiality without requiring cryptography. One lesson to be drawn from these examples is that confidentiality in some cases can be achieved without complex operations.

The first protocol solves an instance of the following problem: how two agents may exchange secret information in the open, without an eavesdropping third agent learning about the information. The instance of the problem, called the Russian Cards problem, is pleasantly concrete and can be explained to children: Alice and Betty each draw three cards from a pack of seven cards, and Eve (the eavesdropper) gets the remaining card. Can players Alice and Betty learn each other's cards without revealing that information to Eve? The restriction is that Alice and Betty can only make public announcements that Eve can hear.

Several protocols for solving the Russian Cards problem have been proposed; a fairly simple solution is the *Seven Hands protocol*. Recall that there are seven cards: three are dealt to Alice, three are dealt to Betty, and the last card is dealt to Eve. Call the cards dealt to Alice  $a_1, a_2, a_3$ , and the cards dealt to Betty  $b_1, b_2, b_3$ . The card dealt to Eve is  $e$ .

The Seven Hands protocol is a two-step protocol that Alice can use to tell her cards to Betty and learn Betty's cards in response:

$$\begin{array}{lll} 1. & A \longrightarrow B & : \quad SH_A \\ 2. & B \longrightarrow A & : \quad SH_B \end{array} \quad (2)$$

where  $SH_A$  and  $SH_B$  are the following specific messages:

- (1) Message  $SH_A$  is constructed by Alice as follows. Alice first chooses a random renaming  $W, X, Y, Z$  of the elements in  $\{b_1, b_2, b_3, e\}$ , that is, a random permutation of the four cards not in her hand. Message  $SH_A$  then consists of the following seven subsets of cards, in some arbitrary order:

$$\begin{array}{c} \{a_1, a_2, a_3\} \\ \{a_1, W, X\} \quad \{a_1, Y, Z\} \\ \{a_2, W, Y\} \quad \{a_2, X, Z\} \\ \{a_3, W, Z\} \quad \{a_3, X, Y\} \end{array}$$

These subsets are carefully chosen: for every possible hand of Betty, that is, for every possible subset  $S$  of size three of  $\{W, X, Y, Z\}$ , there is exactly one set in  $SH_A$  with which  $S$  has an empty intersection, and that set is Alice's hand  $\{a_1, a_2, a_3\}$ . Thus, upon receiving  $SH_A$ , Betty can identify Alice's hand by examining the sets Alice sent and picking the one with which her own hand has an empty intersection, and in the process Betty learns Alice's hand, and by elimination, Eve's card.

- (2) Message  $SH_B$ , Betty's response, is simply Eve's card. Alice, upon receiving  $SH_B$ , knows her own hand and Eve's card, and therefore can infer by elimination Betty's hand.

At the end of the exchange, Alice knows Betty's hand and Betty knows Alice's hand, as required.

What about Eve? She does not learn anything about the cards in Alice or Betty's hand. Indeed, after seeing Alice's message, Eve has no information about Alice's hand, since every card appears in exactly three of the sets Alice sent. There is no way for Eve to isolate

which of those cards might be one of Alice's. Furthermore, after seeing Betty's message, all she has learned is her own card, which she already knew.

If we define  $c \in i$  to be the primitive proposition *card  $c$  is in player  $i$ 's hand* (where  $A, B, E$  represent Alice, Betty, and Eve, respectively) then we expect that the following epistemic formula holds after the first message is received by Betty:

$$K_B(a_1 \in A) \wedge K_B(a_2 \in A) \wedge K_B(a_3 \in A),$$

that the following epistemic formula holds after the second message is received by Alice:

$$K_A(b_1 \in B) \wedge K_A(b_2 \in B) \wedge K_B(b_3 \in B),$$

and that the following formula holds after either of the messages is received:

$$\begin{aligned} &\neg K_E(a_1 \in A) \wedge \neg K_E(a_2 \in A) \wedge \neg K_E(a_3 \in A) \\ &\wedge \neg K_E(b_1 \in B) \wedge \neg K_E(b_2 \in B) \wedge \neg K_E(b_3 \in B). \end{aligned}$$

It is an easy exercise to construct the Kripke structures describing this scenario.

The Seven Hands protocol is ideally suited for epistemic reasoning via a possible-worlds semantics for knowledge, as it relies on combinatorial analysis. Its applicability, however, is limited.

The second protocol is a protocol to ensure anonymity, which is a form of confidentiality (see §4). It does not rely on combinatorial analysis but rather on properties of the XOR operation.<sup>3</sup> The Dining Cryptographers protocol was originally developed to solve the following problem. Suppose that Alice, Betty, and Charlene are three cryptographers having dinner at their favorite restaurant. Their waiter informs them that arrangements have been made for the bill to be paid anonymously by one party. That payer might be one of the cryptographers, but it might also be U.S. National Security Agency. The three cryptographers respect each other's right to make an anonymous payment, but they would like to know whether the NSA is paying.

The following protocol can be used to satisfy the cryptographers' curiosity and allow each of them to determine whether the NSA or one of her colleagues is paying, without revealing the identity of the payer in the latter case.

- (1) Every cryptographer  $i$  flips a fair coin privately with her neighbor  $j$  on her right: the Boolean result  $T_{\{i,j\}}$  is *true* if the coin lands tails, and *false* if the coin lands heads. Thus, the cryptographers produce the Boolean results  $T_{\{A,B\}}$ ,  $T_{\{A,C\}}$ ,  $T_{\{B,C\}}$ ; Alice sees  $T_{\{A,B\}}$  and  $T_{\{A,C\}}$ ; Betty sees  $T_{\{A,B\}}$  and  $T_{\{B,C\}}$ ; Charlene sees  $T_{\{A,C\}}$  and  $T_{\{B,C\}}$ .
- (2) Every cryptographer  $i$  computes a private Boolean value  $Df_i$  as *true* if the two coin tosses she has witnessed are different, and *false* if they are the same. Thus,  $Df_A = T_{\{A,B\}} \oplus T_{\{A,C\}}$ ,  $Df_B = T_{\{A,B\}} \oplus T_{\{B,C\}}$ , and  $Df_C = T_{\{A,C\}} \oplus T_{\{B,C\}}$ .
- (3) Every cryptographer  $i$  publicly announces  $Df_i$ , except for the paying cryptographer (if there is one) who announces  $\neg Df_i$ , the negation of  $Df_i$ .

---

<sup>3</sup>XOR (exclusive or) is a binary Boolean operation  $\oplus$  defined by taking  $b_1 \oplus b_2$  to be true if and only if exactly one of  $b_1$  or  $b_2$  is true. It is associative and commutative.

Once the protocol is executed, any curious cryptographer interested in determining who paid for dinner simply has to take the XOR of all the announcements: if the result is *false*, then the NSA paid, and if the result is *true*, then one of the cryptographers paid.

To see why this is the case, consider the two possible scenarios. Suppose the NSA paid. Then the XOR of all the announcements is:

$$\begin{aligned}
Df_A \oplus Df_B \oplus Df_C &= (T_{\{A,B\}} \oplus T_{\{A,C\}}) \oplus (T_{\{B,C\}} \oplus T_{\{A,B\}}) \oplus (T_{\{A,C\}} \oplus T_{\{B,C\}}) \\
&= (T_{\{A,B\}} \oplus T_{\{A,B\}}) \oplus (T_{\{B,C\}} \oplus T_{\{B,C\}}) \oplus (T_{\{A,C\}} \oplus T_{\{A,C\}}) \\
&= false \oplus false \oplus false \\
&= false
\end{aligned}$$

whereas if one of the cryptographers paid (without loss of generality, suppose it is Alice), then the XOR of all the announcements is:

$$\begin{aligned}
\neg Df_A \oplus Df_B \oplus Df_C &= \neg (T_{\{A,B\}} \oplus T_{\{A,C\}}) \oplus (T_{\{B,C\}} \oplus T_{\{A,B\}}) \oplus (T_{\{A,C\}} \oplus T_{\{B,C\}}) \\
&= (\neg T_{\{A,B\}} \oplus T_{\{A,C\}}) \oplus (T_{\{B,C\}} \oplus T_{\{A,B\}}) \oplus (T_{\{A,C\}} \oplus T_{\{B,C\}}) \\
&= (\neg T_{\{A,B\}} \oplus T_{\{A,B\}}) \oplus (T_{\{B,C\}} \oplus T_{\{B,C\}}) \oplus (T_{\{A,C\}} \oplus T_{\{A,C\}}) \\
&= true \oplus false \oplus false \\
&= true
\end{aligned}$$

If one of the cryptographers paid, neither of the two other cryptographers will know which of her colleagues paid, since either possibility is compatible with what they can observe. Again, it is an easy exercise to construct the Kripke structures capturing these scenarios.

## 2.2 Cryptography

While protocols such as the Seven Hands protocol and the Dining Cryptographers protocol enforce confidentiality by carefully constructing specific messages meant to convey specific information in a specific context, most cryptographic protocols rely on cryptography for confidentiality.

Cryptography seems a natural approach for confidentiality. After all, the whole point of cryptography is to hide information in such a way that only agents with a suitable key can access the information. And indeed, if the goal is for Alice to send message  $m$  to Betty when Alice and Betty alone share a key to encrypt and decrypt messages, then the simplest protocol for confidential message exchange is simply for Alice to encrypt  $m$  and send it to Betty. But how do Alice and Betty come to share a key in the first place? Distributing keys is tricky, because keys have to be sent to the right agents, in such a way that no other agent can get them.

Before addressing those problems, let us review the basics of cryptography. The reader is assumed to have been exposed to at least informal descriptions of cryptography. An encryption scheme is defined by a set of sourcetexts, a set of ciphertexts, a set of keys, and for every key  $k$  an injective encryption function  $e_k$  producing a ciphertext from a sourcetext

and a decryption function  $d_k$  producing a sourcetext from a ciphertext, with the property that  $d_k(e_k(x)) = x$  for all sourcetexts  $x$ . We often assume that ciphertexts and keys are included in sourcetexts to allow for nested encryption and encrypted keys.<sup>4</sup>

There are two broad classes of encryption schemes, which differ in how keys are used for decryption. *Shared-key encryption schemes* require an agent to have a full key to both encrypt and decrypt a message. They tend to be efficient, and can often be implemented directly in hardware. *Public-key encryption schemes*, on the other hand, are set up so that an agent only needs to know part of a key (called the public key) to encrypt a message, while needing the full key to decrypt a message. The full key cannot be easily recovered from knowing only the public key. Public keys are generally made public (hence the name), so that any agent can encrypt a message intended for, say, Alice, by looking up and using Alice's public key. Since only Alice has the full key, only she can decrypt that message. DES and AES are concrete examples of shared-key encryption schemes, while RSA and elliptic-curve encryption schemes are concrete examples of public-key encryption schemes.

Cryptographic protocols are needed with shared-key encryption schemes because agents need to share a key in order to exchange encrypted messages. How is such a shared key distributed? And how can agents make sure they are not tricked into sharing those keys with attackers? An additional difficulty is that when the same shared key is reused for every interaction between two agents, the content of all those interactions becomes available to an attacker that manages to learn that key. To minimize the impact of a key compromise, many systems create a fresh session key for any two agents that want to communicate, which exacerbates the key distribution problem.

Public-key encryption simplifies key distribution, since public keys can simply be published. Any agent wanting to send a confidential message to Alice has only to look up Alice's public key and use it to encrypt her message. The problem, from Alice's perspective, is that anyone can encrypt a message and send it to her, which means that if Alice wants to make sure that the encrypted message she received came from Betty, some sort of authentication mechanism is needed. Furthermore, all known public-key encryption schemes are computationally expensive, so a common approach is to have agents that want to exchange messages in a session first use public-key encryption to generate a session key for a shared-key encryption scheme that they use for their exchanged messages. Such a scenario requires authentication to ensure that agents are not tricked into communicating with an attacker.

**Sample Cryptographic Protocols.** Most classical cryptographic protocols are designed to solve the problem of key distribution for shared-key encryption schemes, and of authentication for public-key encryption schemes. In these contexts, confidentiality and authentication are the key properties: confidentiality to enforce that distributed keys remain secret from attackers, and authentication to ensure that agents can establish the identity of the other agents involved in a message exchange.

This section presents two protocols, each illustrating different problems that can arise and highlighting vulnerabilities that attackers can exploit. (Attackers will be introduced more carefully in the next section.) The first protocol distributes session keys for a shared-

---

<sup>4</sup>This section considers deterministic encryption schemes only, ignoring probabilistic encryption schemes.



key encryption scheme, while the second protocol aims at achieving mutual authentication for public-key encryption schemes. Not all of the problems illustrated will occur in every protocol, of course, nor are vulnerabilities in one context necessarily vulnerabilities in another context.

For the first protocol, consider the following situation. Suppose Alice wants to communicate with Betty, and there is a trusted server Serena who will generate a shared session key (for some shared-key encryption scheme) for them to use. Assume that every registered user of the system shares a distinct key with the trusted server in some shared-key encryption scheme; these keys for Alice and Betty are denoted  $k_{AS}$  and  $k_{BS}$ .

The idea is for Serena to generate a fresh key and send it to both Alice and Betty. Sending it in the clear, however, would allow an eavesdropping attacker to read it and then use it to decrypt messages between Alice and Betty. Since Alice and Betty both share a key with Serena, one solution might be to use those keys to encrypt the session key sent to Alice and Betty, but this turns out to be difficult to implement in practice. Here is the problem. Alice, wanting to communicate with Betty, sends a message to Serena asking her to generate a session key, and Serena sends it to both Alice and Betty. As far as Betty is concerned, she receives a key with an indication that Alice will use it to send her messages. Betty now has to store the key and wait for Alice to send her messages encrypted with that key. If Alice wants to set up several concurrent communications with Betty, then Betty will have to match each incoming communication with the appropriate key, which is annoying at best and inefficient at worst. It turns out to be more efficient for Serena to send the fresh session key  $k_{sess}$  to Alice, and for Alice to forward the key to Betty in her first message. This observation leads to the following protocol:

$$\begin{array}{lll}
1. & A \longrightarrow S & : \quad B \\
2. & S \longrightarrow A & : \quad B, \{k_{sess}\}_{k_{AS}}, \{k_{sess}\}_{k_{BS}} \\
3. & A \longrightarrow B & : \quad A, \{k_{sess}\}_{k_{BS}}
\end{array} \tag{3}$$

Both Alice and Betty learn key  $k_{sess}$ , which is kept secret from eavesdroppers.

While this protocol might seem sufficient to distribute a key to both Alice and Betty, several things can go wrong in the presence of an insider attacker, that is, an attacker that is also a registered user of the system and has control over the network (i.e., can intercept and forge messages; see §2.3).

The insider attacker, Isabel, can initiate a communication with trusted server Serena via her shared key  $k_{IS}$  (Isabel is assumed to have such a key because she is a registered user of the system), and use the key to pose as Alice to Betty. Here is a sequence of messages exemplifying the attack, where the notation  $I[A]$  denotes  $I$  posing as  $A$ :<sup>5</sup>

$$\begin{array}{lll}
I & \longrightarrow S & : \quad B \\
S & \longrightarrow I & : \quad B, \{k_{sess}\}_{k_{IS}}, \{k_{sess}\}_{k_{BS}} \\
I[A] & \longrightarrow B & : \quad A, \{k_{sess}\}_{k_{BS}}
\end{array} \tag{4}$$

Betty believes that she is sharing key  $k_{sess}$  with Alice, while she is in fact sharing it with

---

<sup>5</sup>We can assume that every message has a *from* and *to* field—think email—and that these can be forged. Isabel posing as Alice means that Isabel sends a message and forges the *from* field of the message to hold Alice's name.

Isabel. This is a failure of authentication—the protocol does not authenticate the initiator to the responder.

Isabel can also trick Alice into believing she is talking to Betty, by posing as the server and intercepting messages between Alice and the server, as the following sequence of messages exemplifies:

$$\begin{array}{llll}
A & \longrightarrow & I[S] & : \quad B \\
I[A] & \longrightarrow & S & : \quad I \\
S & \longrightarrow & I[A] & : \quad I, \{k_{sess}\}_{k_{AS}}, \{k_{sess}\}_{k_{IS}} \\
I[S] & \longrightarrow & A & : \quad B, \{k_{sess}\}_{k_{AS}}, \{k_{sess}\}_{k_{IS}} \\
A & \longrightarrow & I[B] & : \quad A, \{k_{sess}\}_{k_{IS}}
\end{array} \tag{5}$$

This form of attack is commonly known as a man-in-the-middle attack. Isabel intercepts Alice’s message to the server, and turns around and sends a different message to the server posing as Alice. The response from the server is intercepted by Isabel, who crafts a suitable response back to Alice. Alice takes that response (which she believes is coming from the server) and sends it to Betty, but that message is intercepted by Isabel as well. Now, as far as Alice is concerned, she has successfully completed the protocol, and holds a key  $k_{IS}$  that she believes she can use to communicate confidentially with Betty, while she is really communicating with Isabel.

How can we correct these vulnerabilities? One feature on which these attacks rely is that the identity of the intended parties for the keys in the protocol are easily forged by the attacker. So one fix is to bind the intended parties to the appropriate copies of the key. Here is an amended version of the protocol:

$$\begin{array}{llll}
1. & A & \longrightarrow & S : B \\
2. & S & \longrightarrow & A : B, \{B, k_{sess}\}_{k_{AS}}, \{A, k_{sess}\}_{k_{BS}} \\
3. & A & \longrightarrow & B : \{A, k_{sess}\}_{k_{BS}}
\end{array} \tag{6}$$

When Alice receives her response from the server and decrypts her message  $\{B, k_{sess}\}_{k_{AS}}$ , she can verify that the key she meant Serena to create to communicate with Betty is in fact a key meant to communicate with Betty. This suffices to foil Isabel in attack (5). Similarly, when Betty receives her message from Alice containing  $\{A, k_{sess}\}_{k_{BS}}$ , she can verify that the key is meant to communicate with Alice. This suffices to foil attack (4).

Protocol (6) now seems to work as intended. It does suffer from another potential vulnerability, though, one that is less directly threatening, but can still cause problems: it is susceptible to a replay attack. Here is the scenario. Suppose that Isabel eavesdrops on messages as Alice gets a session key  $k_0$  from the trusted server to communicate with Betty, and holds on to messages  $\{B, k_0\}_{k_{AS}}$  and  $\{A, k_0\}_{k_{BS}}$ . Suppose further that after a long delay Isabel manages to somehow obtain key  $k_0$ , perhaps by breaking into Alice’s or Betty’s computer, or by expending several months’ worth of effort to crack the encryption. Once Isabel has  $k_0$ , she can subvert an attempt by Alice to get a session key for communicating with Betty by simply intercepting the messages from Alice to Serena, and replaying the messages  $\{B, k_0\}_{k_{AS}}$  and  $\{A, k_0\}_{k_{BS}}$  she intercepted in the past. The following sequence of

messages exemplifies this attack:

$$\begin{array}{lll}
A \longrightarrow I[S] & : & B \\
I[S] \longrightarrow A & : & B, \{B, k_0\}_{k_{AS}}, \{A, k_0\}_{k_{BS}} \\
A \longrightarrow B & : & \{A, k_0\}_{k_{BS}}
\end{array} \tag{7}$$

The main point here is that Alice and Betty after this protocol interaction end up using key  $k_0$  as their session key, but that key is one that Isabel knows, meaning that Isabel can decrypt every single message that Alice and Betty exchange in that session. So even though she does not have the shared keys  $k_{AS}$  and  $k_{BS}$ , she has managed to trick Alice and Betty into using a key she knows.

Preventing this kind of replay attack requires ensuring that messages from earlier executions of the protocol cannot be used in later executions. One way to do that is to have every agent record every message they have ever sent and received, but that is too expensive to be practical. The common alternative is to use timestamps, or nonces. A nonce is a large random number, meant to be unpredictable and essentially unique—the likelihood that the same nonce occurs twice within two different sessions should be negligible. To fix protocol (6) and prevent replay attacks, it suffices for Alice and Betty to generate nonces  $n_A$  and  $n_B$ , respectively, and send them to trusted server Serena so that she can include them in her responses:

$$\begin{array}{lll}
1. & A \longrightarrow B & : \quad n_A \\
2. & B \longrightarrow S & : \quad A, n_A, n_B \\
3. & S \longrightarrow A & : \quad \{B, k_{sess}, n_A\}_{k_{AS}}, \{A, k_{sess}, n_B\}_{k_{BS}} \\
4. & A \longrightarrow B & : \quad \{A, k_{sess}, n_B\}_{k_{BS}}
\end{array} \tag{8}$$

As long as Alice and Betty, when each receives her encrypted message containing the session key, both check that the nonce in the encrypted message is the one that they generated, then they can be confident that the encrypted messages have not been reused from earlier sessions.

Protocol (8) now seems to work as intended and is not vulnerable to replay attacks. But it does not actually guarantee mutual authentication; that is, it does not guarantee to Alice that she is in fact talking to Betty when she believes she is, and to Betty that she is in fact talking to Alice when she believes she is. Consider the following attack, in which attacker Trudy is not an insider—she is not a registered user of the system—but has control of the network and thus can intercept and forge messages. Trudy poses as Betty by intercepting messages from Alice and forging responses:

$$\begin{array}{lll}
A \longrightarrow T[B] & : & n_A \\
T[B] \longrightarrow S & : & A, n_A, n_T \\
S \longrightarrow A & : & \{B, k_{sess}, n_A\}_{k_{AS}}, \{A, k_{sess}, n_T\}_{k_{BS}} \\
A \longrightarrow T[B] & : & \{A, k_{sess}, n_T\}_{k_{BS}}
\end{array} \tag{9}$$

From Alice's perspective, she has completed the protocol by exchanging messages with Betty, and holds a session key for sending confidential messages to Betty. But of course Alice has been communicating with Trudy, and Betty is not even aware of the exchange.

Trudy cannot actually read the messages sent by Alice, so there is no breach of confidentiality, but Trudy has managed to trick Alice into believing she shares a session key with Betty. In terms of knowledge, Alice knows the session key, but she does not know that Betty does.

There is also a way for Trudy to trick Betty into believing she shares a session key with Alice, by posing as Alice:

$$\begin{array}{llll}
T[A] \longrightarrow B & : & n_T \\
B \longrightarrow S & : & A, n_T, n_B \\
S \longrightarrow T[A] & : & \{B, k_{sess}, n_T\}_{k_{AS}}, \{A, k_{sess}, n_B\}_{k_{BS}} \\
T[A] \longrightarrow B & : & \{A, k_{sess}, n_B\}_{k_{BS}}
\end{array} \tag{10}$$

From Betty's perspective, she has completed the protocol by exchanging messages with Alice, and holds a session key for sending confidential messages to Alice. But of course Betty has been communicating with Trudy, and Alice is not even aware of the exchange. In terms of knowledge, Betty knows the session key, but she does not know that Alice does.

Mutual authentication is achieved through an additional nonce exchange at the end of the protocol which uses the newly created session key:

$$\begin{array}{llll}
1. & A \longrightarrow B & : & n_A \\
2. & B \longrightarrow S & : & A, n_A, n_B, n'_B \\
3. & S \longrightarrow A & : & n'_B, \{B, k_{sess}, n_A\}_{k_{AS}}, \{A, k_{sess}, n_B\}_{k_{BS}} \\
4. & A \longrightarrow B & : & n'_A, \{A, k_{sess}, n_B\}_{k_{BS}}, \{A, n'_B\}_{k_{sess}} \\
5. & B \longrightarrow A & : & \{B, n'_A\}_{k_{sess}}
\end{array} \tag{11}$$

This protocol now seems to work as intended without being vulnerable to replay attacks or authentication failures. How can this be guaranteed?

Intuitively, protocol (11) is not susceptible to replay attacks because of the use of nonces: Alice can deduce that the first encrypted component of the third message was not reused from an earlier protocol execution, and Betty can deduce that the first encrypted component of the fourth message was not reused from an earlier protocol execution.

Similarly, mutual authentication in protocol (11) follows from the use of shared keys: Alice can deduce that Serena created the first encrypted component of the third message; and Betty can deduce that Serena created the first encrypted component of the fourth message. Moreover, if Betty believes that  $k_{sess}$  is a key known only to Alice and herself, then she can deduce that Alice created the second encrypted component in the fourth message, and similarly, if Alice believes that  $k_{sess}$  is a key known only to Betty and herself, then she can deduce that Betty created the encrypted component in the fifth message.

The confidentiality of the session key requires the assumption that trusted server Serena is indeed trustworthy and creates keys that have not previously been used and distributed to other parties. If so, then Alice can deduce that the session key she receives in the third message is a confidential key for communicating with Betty, since Alice can also deduce that she has been executing the protocol with Betty. Similarly, Betty can deduce that the session key she receives in the fourth message is a confidential key for communicating with Alice, since Betty can deduce that she has been executing the protocol with Alice.

In a precise sense, the goal of cryptographic protocol analysis is to prove these kind of properties formally, and many techniques have been developed which are surveyed below in §2.5.

The second cryptographic protocol uses a public-key encryption scheme to achieve mutual authentication:

$$\begin{array}{lll}
1. & A \longrightarrow B & : \quad \{A, n_A\}_{pk_B} \\
2. & B \longrightarrow A & : \quad \{n_A, n_B\}_{pk_A} \\
3. & A \longrightarrow B & : \quad \{n_B\}_{pk_B}
\end{array} \tag{12}$$

where  $pk_A$  and  $pk_B$  are the public keys of Alice and Betty, respectively, and  $n_A$  and  $n_B$  are nonces.

Intuitively, when Alice receives her nonce  $n_A$  back in the second message, she knows that Betty must have decrypted her first message at some point during the execution of the protocol (because only Betty could have decrypted the message that contained it), and similarly, when Betty receives her nonce  $n_B$  back, she knows that Alice must have decrypted her second message. Note that  $n_A$  and  $n_B$  are kept confidential throughout the protocol, and that mutual authentication relies on that confidentiality.

Protocol (12), known as the Needham-Schroeder protocol, achieves mutual authentication even in the presence of an attacker that has control of the network and can intercept and forge messages. It is however vulnerable to insider attackers that are registered users of the system and have control of the network. For example, insider attacker Isabel can use an attempt by Alice to initiate an authentication session with her to trick unsuspecting Betty into believing that Alice is initiating an authentication session with her:

$$\begin{array}{lll}
A \longrightarrow I & : & \{A, n_A\}_{pk_I} \\
I[A] \longrightarrow B & : & \{A, n_A\}_{pk_B} \\
B \longrightarrow I[A] & : & \{n_A, n_B\}_{pk_A} \\
I \longrightarrow A & : & \{n_A, n_B\}_{pk_A} \\
A \longrightarrow I & : & \{n_B\}_{pk_I} \\
I[A] \longrightarrow B & : & \{n_B\}_{pk_B}
\end{array} \tag{13}$$

From Alice's perspective, she has managed to complete a mutual authentication session with Isabel, which was her goal all along. But Isabel also managed to complete an authentication session with Betty, tricking Betty into believing she is interacting with Alice.

There is a simple fix that eliminates that vulnerability:

$$\begin{array}{lll}
1. & A \longrightarrow B & : \quad \{A, n_A\}_{pk_B} \\
2. & B \longrightarrow A & : \quad \{B, n_A, n_B\}_{pk_A} \\
3. & A \longrightarrow B & : \quad \{n_B\}_{pk_B}
\end{array} \tag{14}$$

It is interesting to see how the fix works: if Alice, during her mutual authentication attempt with Isabel, notices that the response message she receives from Isabel names a different agent than Isabel, then she can deduce that her authentication attempt is being subverted to try to confound another agent, and she can abort the authentication attempt at that point.

## 2.3 Attackers

A distinguishing feature of cryptographic protocols, besides the use of cryptography, is that they are deployed in potentially hostile environments in which attackers may attempt to subvert the operations of the protocol.

Reasoning about cryptographic protocols, therefore, requires a threat model, describing the kind of attackers against which the cryptographic protocol should protect. Attackers commonly considered in the literature include:

- **Eavesdropping attackers:** assumed to be able to read all messages exchanged between agents. Eavesdropping attackers do not affect communication in any way, however, and remain hidden from other agents.
- **Active attackers:** assumed to have complete control over communications between agents, that is, able to read all messages as well as intercept them and forge new messages. They remain hidden from other agents, and thus no agent will intentionally attempt to communicate with an active attacker.
- **Insider attackers:** assumed to have complete control over communications between agents just like active attackers, but also considered legitimate registered users in their own right. They can therefore initiate interactions with other agents as themselves, and other agents can intentionally initiate interactions with them.<sup>6</sup>

The class of insider attackers includes the class of active attackers, which itself includes the class of eavesdropping attackers. Thus, in that sense, an insider attacker is stronger than an active attacker which is stronger than an eavesdropping attacker. In practice, this means that a cryptographic protocol that is deemed secure in the presence of an insider attacker will remain so in the presence of active and eavesdropping attackers, and so on.

We saw several examples of attacks in §2.2, performed by different kind of attackers. Most of the protocols in §2.2 achieve their goals in the presence of eavesdropping attackers, while some also achieve their goals in the presence of active attackers but fail in the presence of insider attackers. The Needham-Schroeder protocol (12), for instance, can be shown to satisfy mutual authentication in the presence of active attackers, but not in the presence of insider attackers—as exemplified by attack (13)—while the variant protocol (14) achieves mutual authentication even in the presence of insider attackers.

The attacks described in §2.2 took place at the level of the protocols themselves, and not at the level of the encryption schemes used by the protocols. But vulnerabilities in encryption schemes are also relevant: an attacker cracking an encrypted message from the trusted server to the agents in protocol (11) will learn the session key, which will invalidate any confidentiality guarantees claimed for the protocol. Despite this, cryptographic protocols are typically analyzed independently from the details of the encryption scheme. The main reason is that it abstracts away from vulnerabilities specific to the encryption scheme used, leaving only those relating to the cryptographic protocol. Vulnerabilities in encryption schemes are usually independent of the cryptographic protocols that use them,

---

<sup>6</sup>A fourth class of attackers, less commonly considered, shares characteristics with both eavesdropping attackers and insider attackers: dishonest agents are assumed not to have control over the network but may attempt to subvert the protocol while acting within the limits imposed on legitimate users.

and can be investigated separately. A vulnerability in the protocol will be a vulnerability no matter what encryption scheme is used, and requires a change in the protocol to correct the flaw.

The standard way to analyze cryptographic protocols independently of any encryption scheme is to use a *formal model of cryptography* that assumes perfect encryption leaking no information about encrypted content. It can be defined as the following symbolic encryption scheme. If  $P$  is a set of plaintexts and  $K$  is a set of keys, then the set of sourcetexts is taken to be the smallest set  $S$  of symbolic terms containing  $P$  and  $K$  such that  $(x, y) \in S$  and  $\{x\}_k \in S$  when  $x, y \in S$  and  $k \in K$ . Intuitively,  $(x, y)$  represents the concatenation of  $x$  and  $y$ , and  $\{x\}_k$  represents the encryption of  $x$  with key  $k$ . The ciphertexts are all sourcetexts of the form  $\{x\}_k$ . The symbolic encryption function  $e_k(x)$  simply returns  $\{x\}_k$ , and the symbolic decryption function  $d_k(x)$  returns  $y$  if  $x$  is  $\{y\}_k$ , and some special token **fail** otherwise.<sup>7</sup>

In the context of analyzing protocols with a formal model of cryptography, attackers are usually modeled using *Dolev-Yao capabilities*. These capabilities go hand in hand with the symbolic aspect of formal models of cryptography. Intuitively, eavesdropping Dolev-Yao attackers can split up concatenated messages and decrypt them if they know the decryption key; active Dolev-Yao attackers can additionally create new messages by concatenating existing messages and encrypting them with known keys. Dolev-Yao attackers do not have the capability of cracking encryptions, nor can they access messages at the level of their component bits.

## 2.4 Modeling Knowledge

The analyses in §2.2 show that various notions of knowledge arise rather naturally when reasoning informally about properties of cryptographic protocols. There are essentially two main kinds of knowledge described in the literature. In some frameworks, both kinds of knowledge are used.

**Message Knowledge.** The first kind of knowledge, the most common and in some sense the most straightforward, tries to capture the notion of *knowing a message*.

There are several equivalent approaches to modeling this kind of knowledge, at least in a formal model of cryptography with Dolev-Yao capabilities. Intuitively, the idea is a constructive one: an attacker knows a message if she can construct that message from other messages she has received or intercepted. (Message knowledge in the context of confidentiality properties is often presented from the perspective of an attacker, since confidentiality is breached when the attacker comes to know a particular message.) In such a context, knowing a message is sometimes called *having a message*, *possessing a message*, or *seeing a message*.

Message knowledge may be described via the following sets. Let  $H$  be a set of messages that the attacker has received or intercepted. The set  $Parts(H)$ , the set of all components

---

<sup>7</sup>The symbolic decryption function embodies an assumption that encrypted messages have enough redundancy for an agent to determine when decryption is successful.

of messages from  $H$ , is defined inductively by the following inference rules:

$$\frac{m \in H}{m \in \text{Parts}(H)} \quad \frac{\{m\}_k \in \text{Parts}(H)}{m \in \text{Parts}(H)}$$

$$\frac{(m_1, m_2) \in \text{Parts}(H)}{m_1 \in \text{Parts}(H)} \quad \frac{(m_1, m_2) \in \text{Parts}(H)}{m_2 \in \text{Parts}(H)}$$

We see that the content of all encrypted messages in  $H$  is included in  $\text{Parts}(H)$ , even those that the attacker cannot decrypt. In a sense,  $\text{Parts}(H)$  is an upper bound on messages the attacker can know. In contrast, the set  $\text{Analyzed}(H)$  of messages that the attacker can actually see is more restricted:

$$\frac{m \in H}{m \in \text{Analyzed}(H)}$$

$$\frac{(m_1, m_2) \in \text{Analyzed}(H)}{m_1 \in \text{Analyzed}(H)} \quad \frac{(m_1, m_2) \in \text{Analyzed}(H)}{m_2 \in \text{Analyzed}(H)}$$

$$\frac{\{m\}_k \in \text{Analyzed}(H) \quad k \in \text{Analyzed}(H)}{m \in \text{Analyzed}(H)}$$

Clearly,  $\text{Analyzed}(H) \subseteq \text{Parts}(H)$ . One definition of message knowledge is to say that an attacker knows message  $m$  in a state where she has received or intercepted a set  $H$  of messages if  $m \in \text{Analyzed}(H)$ . This is the *attacker knows what she can see* interpretation of message knowledge.

The best way to understand this concept of knowledge is to use a physical analogy: we can think of plaintext messages as stones, and encrypted messages as locked boxes. Encrypting a message means putting it in a box and locking it. A message is known if it can be held in one's hands. An encrypted message is known because the box can be held. The content of an encrypted message is known only if the box can be opened (decrypted) and the content (a stone or another box) taken and held.

This form of message knowledge can be captured fairly easily in any logic without using heavy technical machinery, since the data required to define message knowledge is purely local. If we let  $\text{Messages}_i(s)$  be the set of messages received or intercepted by agent  $i$  in state  $s$  of the system, then we can capture message knowledge via a proposition  $\text{knows}_i(m)$ , where  $i$  is an agent and  $m$  is a message, defined to be true at state  $s$  if and only if  $m \in \text{Analyzed}(\text{Messages}_i(s))$ .

Rather than using a dedicated proposition, another approach relies on a dedicated modal operator to capture message knowledge. Message knowledge as defined above can be seen as a form of *explicit knowledge*, often represented by a modal operator  $X_i\varphi$ , read *agent  $i$  explicitly knows  $\varphi$* . (Explicit knowledge is to be contrasted with the implicit knowledge captured by the possible-worlds definition of knowledge.) One form of explicit knowledge, *algorithmic knowledge*, uses a local algorithm stored in the local state of an agent to determine if  $\varphi$  is explicitly known to that agent. Thus,  $X_i\varphi$  is true at a state  $s$  if the local algorithm of agent  $i$  says that the agent knows  $\varphi$  in state  $s$ . If we let proposition  $\text{part}_i(m)$  be true at a state  $s$  when  $m \in \text{Parts}(\text{Messages}_i(s))$ , then it is a simple matter to define a local algorithm to check if  $m \in \text{Analyzed}(\text{Messages}_i(s))$  and capture knowledge of



message  $m$  via  $X_i(part_i(m))$ : agent  $i$  explicitly knows that message  $m$  is part of the messages she has received. Thus,  $part_i(m)$  may be true at a state while  $X_i(part_i(m))$  is false at that state if the message is encrypted with a key that agent  $i$  does not know.

A variant of the *can see* interpretation of message knowledge is to consider instead the messages that an attacker can create. The set  $Synthesized(H)$  of messages that the attacker can create from a set  $H$  of messages is inductively defined by the following inference rules:

$$\begin{array}{c} \frac{m \in H}{m \in Synthesized(H)} \\[1em] \frac{m_1 \in Synthesized(H) \quad m_2 \in Synthesized(H)}{(m_1, m_2) \in Synthesized(H)} \\[1em] \frac{m \in Synthesized(H) \quad k \in Synthesized(H)}{\{m\}_k \in Synthesized(H)} \end{array}$$

An alternative interpretation of message knowledge, the *attacker knows what she can send* interpretation, can be defined as: an attacker knows message  $m$  in a state where she has received or intercepted a set  $H$  of messages if  $m \in Synthesized(Analyzed(H))$ . Since  $Analyzed(H) \subseteq Synthesized(Analyzed(H))$ , everything an attacker can see she can also send.

The *can send* interpretation of message knowledge is tricky, because clearly any agent can send any plaintext and any key—this is akin to being able to send any password—and it is easy to inadvertently define nondeterministic attackers that can synthesize any message. The intent is for attackers to be able to send only messages based on those she has received or intercepted, but that is a restriction that can be difficult to justify. This suggests some subtleties in choosing the right definition of message knowledge.

Message knowledge, whether under the *can see* or *can send* interpretation, is severely constrained. It is knowledge of terms, as opposed to knowledge of facts—although terms can be facts, facts are more general than terms. Message knowledge is conducive to formal verification using a variety of techniques, mostly because it does not require anything but looking at the local state of an agent. Indeed, message knowledge is inherently local.

**Possible-Worlds Knowledge.** The other kind of knowledge that arises in the study of cryptographic protocols is the standard possible-worlds definition of knowledge via an accessibility relation over the states of a structure. The Kripke structures interpreting knowledge are usually sets of traces of the protocol and the accessibility relation for agent  $i$  is an equivalence relation over the states of the system that relates two states in which agent  $i$  has the same local state (including having received or intercepted the same messages).

In the presence of cryptography, the standard accessibility relation, meant to capture when two states are indistinguishable to an agent, seems inappropriate. After all, the whole point of cryptography is to hide information—and in particular, most cryptographic definitions say that if an agent receives message  $m_1$  encrypted with a key  $k_1$  that she does not know and message  $m_2$  encrypted with key  $k_2$  that she also does not know, then that agent should be unable to distinguish the two messages, in the sense of being able to identify which is which. Thus, goes the argument, a state where an agent has received  $\{m_1\}_{k_1}$

and a state where that agent has received  $\{m_2\}_{k_2}$  instead should be indistinguishable if  $k_1$  and  $k_2$  are not known.

To capture a more appropriate definition of state indistinguishability, one approach is to filter the local states through a function that replaces all messages encrypted with an unknown key by a special token  $\square$ . More precisely, if  $H$  is a set of messages, we write  $[H] = \{[m]^H : m \in H\}$ , where  $[m]^H$  is inductively defined as follows:

$$\begin{aligned} [m]^H &= m && \text{if } m \text{ is a plaintext} \\ [(m_1, m_2)]^H &= ([m_1]^H, [m_2]^H) \\ [\{m\}_k]^H &= \begin{cases} \{[m]^H\}_k & \text{if } k \in \text{Analyzed}(H) \\ \square & \text{otherwise} \end{cases} \end{aligned}$$

The revised equivalence relations  $\sim_i^\square$  through which knowledge is interpreted can now be defined to be  $s \sim_i^\square t$  if and only if  $[Messages_i(s)] = [Messages_i(t)]$ .<sup>8</sup>

The definition  $[-]^H$  above, which is typical, uses  $\text{Analyzed}(-)$  to extract the keys that the agent knows. Alternate definitions can be given, from a simpler definition that looks for keys appearing directly in the local state, to a more complex recursive definition defined using possible-worlds knowledge.

Possible-worlds knowledge interpreted via an  $\sim_i^\square$  accessibility relation is general enough to express message knowledge. If we assume a class of propositions  $part_i(m)$  as before, true at a state  $s$  when  $m \in \text{Parts}(Messages_i(s))$ , then formula  $K_i(part_i(m))$  says that agent  $i$  knows message  $m$ —intuitively, she knows that  $m$  is part of some message in her local state, and has access to it.

To see that  $K_i(part_i(m))$  corresponds to message knowledge as defined above, we can relate it to the definition of message knowledge in terms of a local  $knows_i(m)$  proposition, using the *can see* interpretation of message knowledge. It is not difficult to show that if  $knows_i(m)$  is true at state  $s$ , then  $K_i(part_i(m))$  must also be true at state  $s$ . The converse direction requires a suitable richness condition that guarantees that there are enough encrypted messages to compare: for every message  $\{m\}_k$  received or intercepted where  $k$  is not known to the agent, there should exist another state in which the agent has received  $\{m'\}_{k'}$  for a different  $m'$  and a different  $k'$ .<sup>9</sup> Under such a richness condition, if  $K_i(part_i(m))$  is true at a state  $s$ , then  $knows_i(m)$  is true at that same state  $s$ .

Thus, possible-worlds knowledge can be used to express message knowledge, and can also capture higher-order knowledge, that is, knowledge about general facts, including other agents' knowledge. The informal analyses of §2.2 show that it makes sense to state that Alice may know that Betty knows the key. While Betty's knowledge here is message knowledge (knowledge of the key) and therefore can be modeled with any of the approaches above, Alice's knowledge is higher-order knowledge, knowledge about knowledge of another agent. Logics that allow reasoning about Alice's knowledge of Betty's knowledge of the key tend to rely on possible-worlds definitions of knowledge.

<sup>8</sup>This definition does not account for the possibility that an agent, even if she does not know the content of an encrypted message, may still recognize that she has already seen that encrypted message. (This is an issue when encryption is deterministic, so that encrypting  $m$  with key  $k$  always yields the same string of bits.) One approach is to refine the definition so that every encryption  $\{m\}_k$  is replaced by a unique token  $\square_{m,k}$ .

<sup>9</sup>To see the need for a richness condition, if there is a single state in which agent  $i$  has received an encrypted message, then  $K_i(part_i(m))$  holds vacuously when  $m$  is the content of the encrypted message.

## 2.5 Reasoning about Cryptographic Protocols

Several approaches have been developed for reasoning about cryptographic protocols. Most are not based on epistemic logic, but extend a classical propositional or first-order logic—possibly with temporal operators—with a simple form of message knowledge in the spirit of  $knows_i(m)$ . This allows them to leverage well-understood techniques for system analysis from the formal verification community and from the programming language community. Other approaches are explicitly epistemic in nature.

Techniques for reasoning about cryptographic protocols roughly split along two axes, each corresponding to a way of using logic to reason about protocols in general.

- (1) Reasoning can be performed either deductively using the proof theory of the logic (e.g., through deductions in a theorem prover), or semantically, using the models of the logic (e.g., through model checking).
- (2) Reasoning can be performed either directly on the description of the protocol—either taken as a sequence of messages or a program for each role in the protocol—or indirectly on the set of traces generated by protocol executions.

Comparing reasoning methods across these axes is difficult, as each have their advantages and their disadvantages.

**The Inductive Method.** A good example of a deductive approach for reasoning about security protocols is the Inductive Method, based on inductive definitions in higher-order logic (a generalization of first-order predicate logic allowing quantification over arbitrary relations). These inductive definitions admit powerful induction principles which become the main proof technique used to establish confidentiality and authentication properties.

The Inductive Method is fairly characteristic of many deductive approaches to cryptographic protocol analysis: the deductive system is embedded in a powerful logic such as higher-order logic, and does not use epistemic concepts beyond a local definition of message knowledge equivalent to the use of a  $knows_i(m)$  proposition.

The Inductive Method proper is based on defining a theory—a set of logical rules—for analyzing a given protocol. The theory for a protocol describes how to generate the protocol execution traces, where a trace is a sequence of events such as *A sends m to B*, represented by the predicate  $Say(A, B, m)$ . Rules state which events can possibly follow a given sequence of events, thereby describing traces inductively. In general, there is a rule in the theory for every message in the protocol description. Rules inductively define a set  $Prot$  of traces representing all the possible traces of the protocol.

If we consider a theory for Protocol (14), a rule for message (1) would say:

$$\begin{aligned} tr &\in Prot \\ \Rightarrow \langle tr, Say(A, B, \{A, n_A\}_{pk_B}) \rangle &\in Prot \end{aligned}$$

where  $\langle tr, e \rangle$  adds event  $e$  to trace  $tr$ . That is, if  $tr$  represents a valid trace of the protocol, then that trace can be extended with the first message of a new protocol execution.

Similarly, a rule for message (2) would say:

$$\begin{aligned}
& tr \in \text{Prot} \\
& \wedge \text{Say}(A', B, \{A, n_A\}_{pk_B}) \in tr \\
& \Rightarrow \langle tr, \text{Say}(B, A, \{B, n_A, n_B\}_{pk_A}) \rangle \in \text{Prot}
\end{aligned}$$

That is, if  $tr$  is a valid trace of the protocol in which an agent has received the first message of a protocol execution, that agent can respond appropriately with the second message of the protocol execution.<sup>10</sup> Rules are simply implications and conjunctions over a vocabulary of events.

The attacker  $S$  is also defined by rules; these rules describe how attacker actions can extend traces with new events. For a Dolev-Yao attacker, these rules define a nondeterministic process that can intercept any message, decompose it into parts and decrypt it if the correct key is known, and that can create new messages from other messages it has observed. The theory includes inductive definitions for the *Analyzed* and *Synthesized* sets given in §2.4, as well as rules of the form

$$\begin{aligned}
& tr \in \text{Prot} \\
& \wedge m \in \text{Synthesized}(\text{Analyzed}(\text{Spied}(tr))) \\
& \wedge B \in \text{Agents} \\
& \Rightarrow \langle tr, \text{Say}(S, B, m) \rangle \in \text{Prot}
\end{aligned}$$

that states that if  $m$  can be synthesized from the messages the attacker observed on trace  $tr$  (captured by an inductively-defined set  $\text{Spied}(tr)$ ), then the attacker can add an event  $\text{Say}(S, B, m)$  for any agent  $B$  to the trace.

The Inductive Method is geared for proving safety properties: for every state in every trace, that state is not a bad state. A protocol is proved correct by induction on the length of the traces: choosing the shortest sequence to a bad state, assuming all states earlier on the trace are good, then deriving a contradiction by showing that any state following these good states must be good itself.

A confidentiality property such as *the attacker never learns message  $m$*  is established by making sure that the attacker is unable to ever send message  $m$ , by proving the following formula:

$$(\forall tr \in \text{Prot}) (\forall B \in \text{Agents}) \text{Say}(S, B, m) \notin tr$$

This is a *can send* interpretation of message knowledge. Indeed, according to the rules for the attacker, if the attacker knows message  $m$  at any point during a trace, then there exists an extension of that trace where the attacker sends message  $m$ . Thus, showing that the attacker never learns message  $m$  amounts to showing that there is no trace in which an event  $\text{Say}(S, B, m)$  appears, for any agent  $B$ .

Abstracting away from the details of the approach, the Inductive Method essentially relies on rules to describe the evolution of a protocol execution, and verifying a confidentiality property is reduced to verifying that a certain bad state is not reachable. Other approaches to cryptographic protocol analysis share this methodology, many of them using a

---

<sup>10</sup>These rules are simplifications. Actual rules would contain appropriate quantification and additional side conditions to ensure that  $A$  and  $B$  are different agents, that nonces do not clash, and so on.

logic programming language rather than higher-order logic to express protocol evolution rules; see §6.

**BAN Logic.** The Inductive Method relies on encoding rules for generating protocol execution traces in an expressive general logic suitable for automating inductive proofs. In contrast, the next approach, BAN Logic, is a logic tailored for reasoning about cryptographic protocols described as a sequence of message exchanges. It has the additional feature of including a higher-order *belief* operator as a primitive.

BAN Logic is a logic in the tradition of Hoare Logic, in that it advocates an axiomatic approach for reasoning about cryptographic protocols. BAN Logic tracks the evolution of beliefs during the execution of cryptographic protocol, and is described by a set of inference rules for deriving new beliefs from old. BAN Logic includes primitive formulas stating that  $k$  is a shared key known only to  $A$  and  $B$  ( $A \stackrel{k}{\leftrightarrow} B$ ), that  $m$  is a secret between  $A$  and  $B$  ( $A \stackrel{m}{\equiv} B$ ), that agent  $A$  believes formula  $F$  ( $A$  believes  $F$ ), that agent  $A$  controls the truth of formula  $F$  ( $A$  controls  $F$ ), that agent  $A$  sent a message meaning  $F$  ( $A$  said  $F$ ), that agent  $A$  received and understood a message meaning  $F$  ( $A$  sees  $F$ ), and that a message meaning  $F$  was created during the current protocol execution ( $\text{fresh}(F)$ ). The precise semantics of these formulas is given indirectly through inference rules, some of which are presented below.

BAN Logic assumes that agents can recognize when an encrypted message is one they have created themselves; encryption is in consequence written  $\{F\}_k^i$ , where  $i$  denotes the agent who encrypted a message meaning  $F$  with key  $k$ . (This also highlights another characteristic of BAN Logic: messages are formulas.)

Here are some of the inference rules of BAN Logic:

- |      |  |
|------|--|
| (R1) | $\frac{A \text{ believes } B \stackrel{k}{\leftrightarrow} A \quad A \text{ sees } \{F\}_k^i \quad i \neq A}{A \text{ believes } B \text{ said } F}$ |
| (R2) | $\frac{A \text{ believes } B \text{ said } (F, F')}{A \text{ believes } B \text{ said } F}$  |
| (R3) | $\frac{A \text{ believes } \text{fresh}(F) \quad A \text{ believes } (B \text{ said } F)}{A \text{ believes } B \text{ believes } F}$                |
| (R4) | $\frac{A \text{ believes } B \text{ controls } F \quad A \text{ believes } B \text{ believes } F}{A \text{ believes } F}$                            |
| (R5) | $\frac{A \text{ sees } (F, F')}{A \text{ sees } F}$  |
| (R6) | $\frac{A \text{ believes } B \stackrel{k}{\leftrightarrow} A \quad A \text{ sees } \{F\}_k^i \quad i \neq A}{A \text{ sees } F}$                     |
| (R7) | $\frac{A \text{ believes } \text{fresh}(F)}{A \text{ believes } \text{fresh}((F', F))}$  |
| (R8) | $\frac{A \text{ believes } B \text{ believes } (F, F')}{A \text{ believes } B \text{ believes } F}$  |

Rule (R1), for instance, says that if agent  $A$  believes that  $k$  is shared only between  $B$  and

herself, and she receives a message encrypted with key  $k$  that she did not encrypt herself, then she believes that  $B$  sent the original message. Rule (R3) is an honesty rule: it says that agents send messages meaning  $F$  only when they believe  $F$ . There are commutative variants of rules (R2), (R5), (R7), and (R8), as well as variants for more general tuples; there are also variants of (R8) for any level of nested belief.

BAN Logic does not attempt to model protocol execution traces. Reasoning is done directly on the sequence of messages in the description of the protocol. Because sequences of messages do not carry enough information to permit this kind of reasoning, a transformation known as *idealization* must be applied to the protocol. Roughly speaking, idealization consists of replacing the messages in the protocol by formulas of BAN Logic that capture the intent of each message. For instance, if agent  $A$  sends key  $k$  to agent  $B$  with the intention of sharing a key that is known only to  $A$ , then a suitable idealization would have  $A$  send the formula  $A \xleftrightarrow{k} B$  to  $B$ . Idealization is an annotation mechanism, and as such is somewhat subjective.

To illustrate reasoning in BAN Logic, consider the following simple protocol in which Alice sends a secret value  $m_0$  to Betty encrypted with their shared key  $k_{AB}$ , along with a nonce exchange to convince  $B$  that the message is not a replay of a message in a previous execution of the protocol (see §2.2):

$$\begin{array}{lll} 1. & A \longrightarrow B & : \quad A \\ 2. & B \longrightarrow A & : \quad n_B \\ 3. & A \longrightarrow B & : \quad A, \{m_0, n_B\}_{k_{AB}} \end{array} \quad (15)$$

A possible idealization of protocol (15) would be:

$$3'. \quad A \longrightarrow B : \quad \{A \stackrel{m_0}{\rightleftharpoons} B, n_B\}_{k_{AB}} \quad (16)$$

The first two messages in protocol (15) carry information that BAN Logic does not use, so they are not present in the idealized protocol. The third message is idealized to  $A$  sending formula  $A \stackrel{m_0}{\rightleftharpoons} B$  to  $B$  along with the nonce  $n_B$ , indicating that  $A$  considers  $m_0$  to be a secret at that point.

Reasoning about an idealized protocol consists in laying out the initial beliefs of the agents, and deriving new beliefs from those and from the messages exchanged between the agents, using the inference rules of the logic. For protocol (16), initial beliefs include that both parties believe that key  $k_{AB}$  has not been compromised, that nonce  $n_B$  has not already been used, and that message  $m_0$  that  $A$  wants to send to  $B$  is initially secret. These initial beliefs are captured by the following formulas:

$$\begin{array}{l} A \text{ believes } A \xleftrightarrow{k_{AB}} B \\ B \text{ believes } A \xleftrightarrow{k_{AB}} B \\ B \text{ believes fresh}(n_B) \\ A \text{ believes } A \stackrel{m_0}{\rightleftharpoons} B \end{array}$$

We can derive new formulas from these initial beliefs in combination with the messages exchanged by the agents. The idea is to update this set of formulas after each protocol step:

after an idealized step  $A \rightarrow B : F$ , which says that  $B$  receives a message meaning  $F$ , we can add formula  $B$  sees  $F$  to the set of formulas, and we use the inference rules to derive additional formulas to add to the set.

For example, in idealized protocol (16), after message (3'), we add formula

$$B \text{ sees } \{A \stackrel{m_0}{\rightleftharpoons} B, n_B\}_{k_{AB}} \quad (17)$$

to the set of initial beliefs. Along with the initial belief  $B$  believes  $A \stackrel{k_{AB}}{\rightleftharpoons} B$ , formula (17) allows us to apply inference rule (R1) to derive:

$$B \text{ believes } A \text{ said } (A \stackrel{m_0}{\rightleftharpoons} B, n_B). \quad (18)$$

From the initial belief  $B$  believes fresh( $n_B$ ), inference rule (R7) lets us derive that any message combined with  $n_B$  must be fresh, and thus we can derive:

$$B \text{ believes fresh}(A \stackrel{m_0}{\rightleftharpoons} B, n_B). \quad (19)$$

Formula (19) together with (18) give us, via inference rule (R3):

$$B \text{ believes } A \text{ believes } (A \stackrel{m_0}{\rightleftharpoons} B, n_B).$$

Via inference rule (R8), this yields:

$$B \text{ believes } A \text{ believes } A \stackrel{m_0}{\rightleftharpoons} B. \quad (20)$$

Thus, after the messages of the idealized protocol have been exchanged,  $B$  believes that  $A$  believes that  $m_0$  is a secret between  $A$  and  $B$ . This is about as much as we can expect.

We can say more if we are willing to assume that  $B$  believes that the secrecy of  $m_0$  is in fact controlled by  $A$ . If so, we can add the following formula to the set of initial beliefs:

$$B \text{ believes } A \text{ controls } A \stackrel{m_0}{\rightleftharpoons} B \quad (21)$$

and formulas (21) and (20) combine via inference rule (R4) to yield the stronger conclusion:

$$B \text{ believes } A \stackrel{m_0}{\rightleftharpoons} B.$$

In other words, if  $B$  believes that  $A$  controls the secrecy of  $m_0$  and also that  $A$  believes  $m_0$  to be secret, then after the protocol executes  $B$  also believes that  $m_0$  is a secret shared only with  $A$ .

Attackers are not explicit in BAN Logic. In a sense, an active Dolev-Yao attacker is implicitly encoded within the inference rules of the logic, but the focus of BAN Logic is reasoning about the belief of agents in the presence of an active attacker, as opposed to reasoning about the knowledge of an attacker. A successful attack in BAN Logic shows up as a failure to establish a desired belief for one of the agents following a protocol execution.

**Temporal and Epistemic Temporal Logics.** Another class of approaches for reasoning about cryptographic protocols rely on a form of temporal logic to express desired properties of the protocol and show that they are true of a model representing the protocol—generally through a suitable representation of its traces. This is done through model-checking techniques to determine algorithmically whether a formula is true in the models representing the protocol. These model-checking techniques vary in terms of how the models are described: these can be either directly expressed by finite state machines, or through domain-specific languages.

The simplest approaches to cryptographic protocol analysis via temporal logics merely extend existing temporal-logic verification techniques. At least two challenges arise in these cases: modeling attackers, and expressing message knowledge. For attackers, while eavesdropping attackers do not affect the execution of protocols and therefore are comparatively easy to handle in standard temporal-logic verification frameworks, active attackers require work. In some cases, it is possible to simply encode an active attacker within the model using the tools of the framework. Message knowledge is usually dealt with by introducing a variant of a  $knows_i(m)$  proposition.

In general, the logics themselves are completely straightforward: they are standard propositional or first-order temporal logics extended with a message knowledge predicate. All the action is in the interpretation of the message knowledge predicate, and in the construction of the models to account for the actions of active attackers. There is not much to say about those approaches as far as pertains to epistemic concepts, but they are popular in practice.

More interesting from an epistemic perspective are those frameworks relying on a temporal epistemic logic, that is, a logic with both temporal and epistemic operators. The MCK model-checker is an example of a verification framework that uses a linear-time temporal logic with epistemic operators to verify protocols that do not use cryptography, such as the Seven Hands or the Dining Cryptographers protocols of §2.1. Protocols are described via finite state machines, and formulas express properties of paths through that finite state machine, each such path corresponding to a possible execution of the protocol.

As an example, consider the Dining Cryptographers protocol, which translates well to a finite state machine. States can be described using three agent-indexed Boolean variables  $paid[i]$ ,  $chan[i]$ , and  $df[i]$ , where variable  $paid[i]$  records whether agent  $i$  paid;, variable  $chan[i]$  is a communication channel used by agent  $i$  to send the result of its coin toss to her right neighbor, and variable  $df[i]$  records the announcement of  $Df_i$  by agent  $i$  at the end of the protocol. The initial states are all the states satisfying:

$$(\neg paid[1] \wedge \neg paid[2] \wedge \neg paid[3]) \vee (paid[1] \wedge \neg paid[2] \wedge \neg paid[3]) \\ \vee (\neg paid[1] \wedge paid[2] \wedge \neg paid[3]) \vee (\neg paid[1] \wedge \neg paid[2] \wedge paid[3])$$

Every agent executes the following program, where a single step of the program for each agent is executed in a transition of the state machine:

```
protocol diningcrypto (paid : observable Bool,
                      chan_left, chan_right : Bool,
                      df : observable Bool[])
```



```

coin_left, coin_right : observable Bool

begin
  if    True -> coin_right := True
    [] True -> coin_right := False
  fi;
  chan_right.send(coin_right);
  coin_left := chan_left.recv();
  df[self] := coin_left xor coin_right xor paid;
end

```

Program `diningcrypto`<sup>11</sup> is instantiated for every agent with suitable variables for the parameters:

```

agent 1 executes diningcrypto (paid[1], chan[1], chan[2], df)
agent 2 executes diningcrypto (paid[2], chan[2], chan[3], df)
agent 3 executes diningcrypto (paid[3], chan[3], chan[1], df)

```

At the first state transition, every agent nondeterministically chooses a value for their coin toss into local variable `coin_right`; at the second state transition, the result of the coin toss is sent on the channel given as the `chan_right` parameter; at the third state transition, the local variable `coin_left` for each agent is updated to reflect the result of the coin toss received from the agent's left neighbor; at the fourth state transition, variable `df` is updated for every agent.

Given such a state machine, a formula expressing the anonymity of the payer from agent 1's perspective can be written as:

$$\begin{aligned}
& \mathbf{X}^4 (\neg \text{paid}[1]) \\
& \Rightarrow (K_1(\neg \text{paid}[1] \wedge \neg \text{paid}[2] \wedge \neg \text{paid}[3])) \\
& \quad \vee (K_1(\text{paid}[2] \vee \text{paid}[3]) \wedge \neg K_1 \text{paid}[2] \wedge \neg K_1 \text{paid}[3])
\end{aligned}$$

where  $\mathbf{X}^4$  is a temporal operator meaning *after four rounds*. This formula, which is true or false of an initial state, says that after the protocol terminates, if cryptographer 1 did not pay, then she either knows that no cryptographer paid, or she knows that one of the other two cryptographers paid but does not know which. (Formulas expressing anonymity from agent 2 and agent 3's perspectives are similar.)

MCK has no built-in support for active attackers, so it cannot easily deal with cryptographic protocols even if we were to add a message knowledge primitive to the language that can deal with encrypted messages. Of course, it is possible to encode some attackers within the language that MCK provides for describing models, but the effect on the efficiency of model checking is unclear.

---

<sup>11</sup>The observable annotation is used to derive the indistinguishability relation: two states are indistinguishable to agent  $i$  if the observable variables of the program executed by agent  $i$  have the same value in both states. The `if ... [] ... fi` construct nondeterministically executes one of its branches with an associated condition that evaluates to true. Variable `self` is assigned the name of the agent executing the program.

The theoretical underpinnings of model checking for temporal epistemic logic are fairly well understood, even though the problem has not been studied nearly as much as model checking for temporal logics. Message knowledge does not particularly complicate matters, once the choice of how to interpret message knowledge is made. Accounting for active attackers is more of an issue, since active attackers introduce additional actions into the model, increasing its size.

The main difficulty with model checking epistemic temporal logic is its inherent complexity. While model checking a standard epistemic logic such as S5 takes time polynomial in the size of the model, adding temporal operators and interpreting the logic over the possibly infinite paths in a finite state machine increases that complexity. For example, in the presence of perfect recall (when agents remember their full history) and synchrony (when agents have access to a global clock), the model-checking problem has non-elementary complexity if the logic includes an *until* temporal operator, and is PSPACE-complete otherwise. The problem tends to become PSPACE-complete when perfect recall is dropped. Progress has been made to control the complexity of model-checking epistemic temporal logics by a careful analysis of the complexity of specific classes of formulas that, while restricted, are still sufficiently expressive to capture interesting security properties, but much work remains to be done to make the resulting techniques efficient.

### 3 Information Flow in Multi-Level Systems

Confidentiality in cryptographic protocols is mainly viewed through the lens of access control: some privilege (a key) is required in order to access the confidential data (the content of an encrypted message). An agent who has the key can access the content, an agent without the key cannot. Those access restrictions can control the release of information, but once that information is released there is nothing stopping it from being propagated by agents or by the system through error or malice, or because the released information is needed for the purpose of computations. For systems in which confidentiality is paramount, it is not sufficient to simply ensure that access to confidential data is controlled, there also needs to be a guarantee that even when the confidential data is released it does not land in unauthorized hands. These sorts of confidentiality guarantees require understanding the flow of information in a system.

Confidentiality in the presence of released information is usually studied in the context of systems in which all data are classified with a security level, and where agents have security clearances allowing them to access data at their security level or lower. For simplicity, only scenarios with security levels *high* and *low* (think *classified* and *unclassified* in military settings) will be considered. Intuitively, a high-security agent should be allowed to read both high- and low-security data, and a low-security agent should be allowed to read only low-security data. This is an example of security policy, which describes the forms of information flows that are allowed and those that are disallowed. Information flows that are disallowed capture the desired confidentiality guarantee.

As an example, imagine a commercial system such as a bank mainframe, where agents perform transactions via credit cards or online accounts. In such a system, credit card numbers and bank account numbers might be considered high-security data, and low-security

agents should be prevented from accessing them. However, what about the last four digits of a credit card number? Even that information is often considered sensitive. What about a single digit? What about the digits frequency in any given credit card number? Because it is in general difficult to characterize exactly what kind of information about high-security data should not be leaked to low-security agents, it is often easier to prevent any kind of partial information disclosure.

The problem of preventing information disclosure is made more interesting, and more complicated, by the fact that information may not only flow directly from one point to another (e.g. by an agent sending a message to another, or by information being posted, or by updating an observable memory location) but may also flow indirectly from one point to another. Suppose that the commercial system described above sends an email to a central location whenever a transfer of more than one million dollars into a given account *A* occurs. Anyone observing email traffic can see those emails being sent and learn that account *A* now contains at least a million dollars. This is an extreme example, but it illustrates indirect information flow: information is gained not by directly observing an event, but by correlating an observation with the event.

Epistemic concepts arise naturally in this setting—a security policy saying that there is no flow of information from high-security data to low-security agents can be expressed as *low-security agents do not learn anything about high-security data*. Moreover, the definitions used in the literature essentially rely on a possible-worlds definition of knowledge within a specific class of models.

Two distinct models of information flow will be described. Both of these models are observational models: they define the kind of observations that agents can make about the system and about the activity of other agents. These observations form the basis of agents' knowledge.

The first model considered takes a fairly abstract view of a system, as sequences of events such as inputs from agents, outputs to agents, internal computation, and so on. These events are the observations that agents can make. In such a setting, security policies regulate information about the occurrence of events. The second model considered is more concrete, and stems from practical work on defining verification techniques for information flow at the level of the source code implementing a system. In that model, observations take the form of content of memory locations that programs can manipulate.

### 3.1 Information Flow in Event Systems

The first model of information flow uses sets of traces corresponding to the possible executions of the system. Every trace is a sequence of events; some of those events are high-security events (and only observable by high-security agents), and some of those events are low-security events (and observable by both high-security and low-security agents). The intuition is that a low-security agent, observing only the low-security events in a trace, should not be able to infer any information about the high-security events in a trace.

How can a low-security agent infer information? If we assume that the full set of traces of the system is known to all agents, then a low-security agent, upon observing a particular sequence of low-security events, can narrow down a set of possible traces that could be the actual trace by considering all the traces that are compatible with her view of the low-

security events. By looking at those possible traces, she may infer information about high-security events. For instance, maybe a particular high-security event  $e$  appears in every such possible trace, and thus she learns that high-security event  $e$  has occurred. In the most extreme case, there may be a single trace compatible with her view of the low-security events, and therefore that low-security agent learns exactly which high-security events have occurred.

The model can be formalized using event systems. An event system is a tuple  $S = (E, I, O, Tr)$  where  $E$  is a set of events,  $I \subseteq E$  a set of input events,  $O \subseteq E$  a set of output events, and  $Tr \subseteq E^*$  a set of finite traces representing the possible executions of the system. Given a trace  $\tau \in E^*$  and a subset  $E' \subseteq E$  of events, we write  $\tau|_{E'}$  for the subtrace of  $\tau$  consisting of events from  $E'$  only.

We assign a security level to events in  $E$  by partitioning them into low-security events  $L$  and high-security events  $H$ : events in  $I \cap L$  are low-security input events, events in  $O \cap L$  are low-security output events, and so on.

A naive attempt at defining information flow in this setting might be to say that there is information flowing from high-security events to low-security agents if a low-security agent's view of  $\tau|_L$  implies that at least one high-security event subsequence is not possible. In other words, seeing a particular sequence of low-security events rules out one possible high-security event subsequence. Formally, if we write  $Tr|_H$  for  $\{\tau|_H : \tau \in Tr\}$ , information flows from high-security events to low-security agents if there is a trace  $\tau \in Tr$  such that  $\{\tau'|_H : \tau'|_L = \tau|_L\} \neq Tr|_H$ .

Such a definition turns out to be too strong—it is equivalent to separability described below—because it pinpoints information flows where there are none: since low-security events may influence high-security events, a particular subsequence of high-security events may be ruled out due to the influence of low-security events, and in that case there should be no information flow since the low-security agent could have already predicted that the high-security subsequence would have been ruled out. Intuitively, there is information flow when one high-security event subsequence that should be possible as far as the low-security agent expects is not in fact possible. This argument gives an inkling as to why the definition of information flow is not entirely trivial.

Security policies in event systems are often defined as closure properties of the set of traces. Security policies that historically were deemed interesting for the purpose of formalizing existing multi-level systems include the following:

- **Separability:** no nontrivial interaction between high-security events and low-security agents is possible because for any such interaction there is a trace with the same high-security events but different low-security events, and a trace with the same low-security events but different high-security events. Formally, for every pair of traces  $\tau_1, \tau_2 \in Tr$ , there is a trace  $\tau \in Tr$  such that  $\tau|_L = \tau_1|_L$  and  $\tau|_H = \tau_2|_H$ .
- **Noninference:** a low-security agent cannot learn about the occurrence of high-security events because any trace, as far as the low-security agent can tell, could be a trace where there are no high-security events at all. Formally, for all traces  $\tau \in Tr$ , there is a trace  $\tau' \in Tr$  such that  $\tau|_L = \tau'|_L$  and  $\tau'|_H = \langle \rangle$ .
- **Generalized Noninference:** A more lenient form of noninference, where a low-

security agent cannot learn about the occurrence of high-security input events because any trace could be a trace where there are no high-security input events at all. Formally, for all traces  $\tau \in Tr$ , there is a trace  $\tau' \in Tr$  such that  $\tau|_L = \tau'|_L$  and  $\tau'|_{(H \cap I)} = \langle \rangle$ .

- **Generalized Noninterference:** a low-security agent cannot learn about high-security input events, and high-security input events cannot influence low-security events. Formally, for all traces  $\tau \in Tr$  and all traces  $\tau' \in \text{interleave}((H \cap I)^*, \{\tau|_L\})$ , there is a  $\tau'' \in Tr$  such that  $\tau''|_L = \tau|_L$  and  $\tau''|_{(L \cup (H \cap I))} = \tau'$ . (Function  $\text{interleave}(T, U)$  returns every possible interleaving of every trace from  $T$  with every trace from  $U$ .)

A closure property says that if some traces are in the model, then other variations on these traces must also be in the model. This is clearly an epistemic property. Under a possible-worlds definition of knowledge, an agent knows a formula if that formula is true at all traces that the agent considers possible given her view of the system. In general, the fewer possible traces there are, the more facts can be known, since it is easier for a fact to be true at all possible traces if there are few of them. The closure properties ensure that there are enough possible traces from the perspective of a low-security agent to prevent a specific of class facts from being known.<sup>12</sup>

Closure conditions on sets of traces are therefore just a way to enforce lack of knowledge, given a possible-worlds definition of knowledge. We can make this precise by viewing event systems as Kripke frames.

The accessibility relation of each agent depends on the agent's security level. In the case of interest, a low-security agent is assigned an accessibility relation  $\sim_L$  defined as  $\tau_1 \sim_L \tau_2$  if and only if  $\tau_1|_L = \tau_2|_L$ .

We identify a proposition with a set of traces, intuitively, those traces in which the proposition is true. As usual, conjunction is intersection of propositions, disjunction is union of propositions, negation is complementation of propositions with respect to the full set of traces in the event system, and implication is subset inclusion. To define the proposition *the low-security agent knows P*, we first define the low-security agent's knowledge set of a trace  $\tau$  as the set of all traces  $\sim_L$ -equivalent to  $\tau$ ,  $\mathcal{K}_L(\tau) = \{\tau' : \tau \sim_L \tau'\}$ . The proposition *the low-security agent knows P* can be defined in the usual way, as:

$$\mathcal{K}_L(P) = \{\tau : \mathcal{K}_L(\tau) \subseteq P\}$$

It is easy to see that  $\mathcal{K}_L$  satisfies the usual S5 axioms, suitably modified to account for propositions being sets:

$$\begin{aligned} \text{(D)} \quad & \mathcal{K}_L(P) \cap \mathcal{K}_L(Q) = \mathcal{K}_L(P \cap Q) \\ \text{(K)} \quad & \mathcal{K}_L(P) \subseteq P \\ \text{(PI)} \quad & \mathcal{K}_L(P) \subseteq \mathcal{K}_L(\mathcal{K}_L(P)) \\ \text{(NI)} \quad & \neg \mathcal{K}_L(P) \subseteq \mathcal{K}_L(\neg \mathcal{K}_L(P)) \end{aligned}$$

These properties are the set-theoretic analogues of *Distribution*, *Knowledge*, *Positive Introspection*, and *Negative Introspection*, respectively.

<sup>12</sup>As in §2, this does not take probabilistic information into account.

As an example, consider an event system  $(E, I, O, Tr)$  that satisfies generalized noninterference, and the proposition *high-security input event  $e$  has occurred*. This proposition is represented by the set  $P_e$  of all traces in  $Tr$  in which  $e$  occurs. The proposition *the low-security agent knows that  $e$  has occurred* corresponds to the set of traces  $K_L(P_e)$ . It is easy to check that because the system satisfies generalized noninterference, the set  $K_L(P_e)$  is empty, meaning that there is no trace on which the low-security agent ever knows  $P_e$ , that is, that  $e$  has occurred. By way of contradiction, suppose that  $\tau \in K_L(P_e)$ . By definition,  $\tau \in K_L(P_e)$  if and only if  $K_L(\tau) \subseteq P_e$ . But the closure condition for generalized noninterference implies that there must exist a trace  $\tau' \sim_L \tau$ , that is, a trace in  $K_L(\tau)$ , such that  $e$  does not occur in  $\tau'$ . Thus, there is a trace in  $K_L(\tau)$  which is not in  $P_e$ , and  $K_L(\tau) \not\subseteq P_e$ . Thus,  $\tau \notin K_L(P_e)$ , a contradiction.

This is a somewhat roundabout way to see that there is an implicit epistemic logic lurking which explains the notions of information flow security policies in event systems. It is certainly possible to make such a logic explicit by introducing a syntax and adding an interpretation to event systems, and study information flow in event systems from such a perspective.

The key point here is that event-system models of information flow and the expression of security policies in those models intrinsically use epistemic concepts, and all reasoning is essentially classical epistemic reasoning performed directly on the models.

### 3.2 Language-Based Noninterference

A more concrete model for information flow is obtained by moving away from trace-based models of systems and relying instead on the program code implementing those systems.

Defining information flow at the level of programs has several advantages: the system is described in detail, information can be defined in terms of the data explicitly manipulated by the program, and enforcement can be automated; the latter turns out to be especially important given the complexity of modern computing systems which makes manual analysis often infeasible.

The observational model used by most language-based information-flow security research is not event-based, although it is still broadly concerned with input and output. The focus here is on information flow in imperative programs, which operates by changing the state of the environment as a program executes. The state of the environment is represented by a store holding values associated with variables. Variables can be read and written by programs. Every variable is tagged with a security level, describing the security level of the data it contains. A low-security agent can observe all low-security variables, but not the high-security ones. Inputs to programs are modeled as initial values of variables, while outputs are modeled as final value of variables: low-security inputs are initial value of low-security variables, and so on. The basic security policy generally considered is a form of noninterference: that low-security outputs do not reveal anything about high-security inputs, and that high-security inputs do not influence the value of low-security outputs.

Consider the following short programs:

- (P1)  $h := l + 1$
- (P2)  $l := h + 1$
- (P3) if  $l = 0$  then  $h := h + 1$  else  $l := l + 1$
- (P4) if  $h = 0$  then  $h := h + 1$  else  $l := l + 1$

In all of these programs, variable  $h$  is a high-security variable, and variable  $l$  is a low-security variable.

A program executes in a store assigning initial values to variables, and execution steps modify the store until the program terminates in a final store. Several simplifying assumption are made: programs are deterministic, and programs always terminate. This is purely to keep the discussion and the technical machinery light. These restrictions can be lifted easily. Moreover, the programming language under consideration will not be described in detail; the sample programs should be intuitive enough.

How do we formalize noninterference in this setting? A store  $\sigma$  is a mapping from variables  $x$  to values  $\sigma(x)$ . We assume every variable  $x$  is tagged with a security level  $sec(x) \in \{L, H\}$ . Let  $\Sigma$  be the set of all possible stores. We model execution of a program  $C$  using a function  $\llbracket C \rrbracket : \Sigma \rightarrow \Sigma$  from initial stores to final stores. Thus, executing program  $C$  in store  $\sigma$  yields a final store  $\llbracket C \rrbracket(\sigma)$ . For example, executing program (P1) in store  $\langle l \mapsto 5, h \mapsto 10 \rangle$  yields store  $\langle l \mapsto 5, h \mapsto 6 \rangle$ , and executing program (P3) in store  $\langle l \mapsto 5, h \mapsto 10 \rangle$  yields store  $\langle l \mapsto 6, h \mapsto 10 \rangle$ .

Two stores  $\sigma_1$  and  $\sigma_2$  are  $L$ -equivalent, written  $\sigma_1 \approx_L \sigma_2$ , if they assign the same values to the same low-security variables:  $\sigma_1 \approx_L \sigma_2$  if and only if for all variables  $x$  with  $sec(x) = L$ ,  $\sigma_1(x) = \sigma_2(x)$ . A program  $C$  satisfies noninterference if executing  $C$  in two  $L$ -equivalent states (that is, in two states that a low-security agent cannot distinguish) yields two  $L$ -equivalent states: for all  $\sigma_1$  and  $\sigma_2$ , if  $\sigma_1 \approx_L \sigma_2$ , then  $\llbracket C \rrbracket(\sigma_1) \approx_L \llbracket C \rrbracket(\sigma_2)$ .<sup>13</sup>

How do programs (P1–4) fare under this definition of noninterference? Program (P1) clearly satisfies noninterference, since the final value of low-security variable  $l$  does not depend on the value of any high-level variable, while program (P2) clearly does not. The other two programs are more interesting. The final value of low-security variable  $l$  in program (P3) only depends on the initial value of  $l$ , and thus we expect (P3) to satisfy noninterference, and it does. Program (P4), however, does not, as we can see by executing the program in stores  $\langle l \mapsto 0, h \mapsto 0 \rangle$  and  $\langle l \mapsto 0, h \mapsto 1 \rangle$ , both  $\approx_L$ -equivalent, but which yield stores  $\langle l \mapsto 0, h \mapsto 1 \rangle$  and  $\langle l \mapsto 1, h \mapsto 1 \rangle$ , respectively, two stores that cannot be  $L$ -equivalent since they differ in the value they assign to variable  $l$ . And indeed, observing the final value of  $l$  reveals information about the initial value of  $h$ .

Noninterference is usually established by a static analysis of the program code, which approximates the flow of information through a program before execution. While the details of the static analyses are interesting in their own right, they have little to do with epistemic logic beyond providing an approach to verifying a specific kind of epistemic property in a specific context.

<sup>13</sup>Another way of understanding this definition is that it requires the relation on stores induced by program execution to be a refinement of  $L$ -equivalence  $\approx_L$ . If we define  $\llbracket C \rrbracket_{\approx_L}$  as the relation  $\{(\llbracket C \rrbracket(\sigma_1), \llbracket C \rrbracket(\sigma_2)) : \sigma_1 \approx_L \sigma_2\}$ , then the noninterference condition can be rephrased as  $\llbracket C \rrbracket_{\approx_L} \subseteq \approx_L$ .

Recent work on language-based information-flow security has highlighted the practical importance of declassification, that is, the controlled release of high-security data to low-security agents. The problem of password-based authentication illustrates the need for such release: when a low-security agent tries to authenticate herself as a high-security agent, she may be presented with a login screen asking for the password of the high-security agent. That password should of course be considered high-security information. However, the login screen leaks information, since entering an incorrect password will reveal that the attempted password is not the right password, thereby leaking a small amount of information about the correct password. The leak is small, but it exists, and because of it the login screen does not satisfy the above definition of noninterference. Defining a suitable notion of security policy that allows such small release of information while still preventing more important information flow is a complex problem.

While the concepts underlying information-flow security are clearly epistemic in nature—taking stores as possible worlds and  $L$ -equivalence as an accessibility relation for low-security agents—there is no real demand for an explicit epistemic logic in which to describe policies. One reason is that it is in general difficult to precisely nail down, in a given system, what high-security information should be kept from low-security agents. It is simply easier to ask that no information be leaked to low-security agents. This *no information* condition is easier to state semantically than through an explicit logical language—not learning any information in the sense of noninterference can be stated straightforwardly as a relationship between equivalence relations, while if we were to use an epistemic logic, we would have to say something along the lines of *for all formulas  $\varphi$  that do not depend only on the state of the low-security agent,  $\neg K_L \varphi$*  where  $K_L$  expresses the knowledge of that low-security agent. The latter is patently clunkier to work with. It may be the case that an explicit epistemic logic would be more useful in the context of declassification, where not all information needs to be kept from low-security agents.

## 4 Beyond Confidentiality

The focus of this chapter has been on confidentiality, because it is by far the most studied security property. It is not only important, it also underpins several other security properties. Other related properties are also relevant.

**Anonymity.** A specific form of confidentiality is anonymity, where the information to be kept secret is the association between actions and agents who perform them. Anonymity has been studied using epistemic logic, and several related definitions have been proposed and debated.

To discuss anonymity, we need to be able to talk about actions and agents who perform them. Let  $\delta(i, a)$  be a proposition interpreted as *agent  $i$  performed action  $a$* .

The simplest definition of anonymity is lack of knowledge: action  $a$  performed by agent  $i$  is minimally anonymous with respect to agent  $j$  if agent  $j$  does not know that agent  $i$  performed  $a$ . This can be captured by the formula

$$\neg K_j \delta(i, a).$$



Minimal anonymity is, well, minimal. It does not rule out that agent  $j$  may narrow down the list of possible agents that performed  $a$  to agent  $i$  and one other agent. Stronger forms of anonymity can be defined: action  $a$  performed by agent  $i$  is totally anonymous with respect to agent  $j$  if, as far as agent  $j$  is concerned, action  $a$  could have been performed by any agent in the system (except for agent  $j$ ). This can be captured by the formula

$$\delta(i, a) \Rightarrow \bigwedge_{i' \neq j} P_j \delta(i', a)$$

where  $P_i \varphi$  is the usual dual to knowledge,  $\neg K_i(\neg \varphi)$ , read as *agent  $i$  considers  $\varphi$  possible*.

Total anonymity is at the other extreme on the spectrum from minimal anonymity; it is a very strong requirement. Intermediate definitions can be obtained by requiring that actions be anonymous only up to a given set of agents—sometimes called an anonymity set: action  $a$  performed by agent  $i$  is anonymous up to  $I$  with respect to agent  $j$  if, as far as agent  $j$  is concerned, action  $a$  could have been performed by any agent in  $I$ . This can be captured by the formula:

$$\delta(i, a) \Rightarrow \bigwedge_{i' \in I} P_j \delta(i', a).$$

As an example of this last definition of anonymity, note that it can be used to describe the anonymity provided by the Dining Cryptographers protocol from §2.1. Recall that if one of the cryptographers paid, the Dining Cryptographers protocol guarantees that each of the non-paying cryptographers think it possible that any of the cryptographers but herself paid. In other words, if  $C = \{Alice, Betty, Charlene\}$  are the cryptographers and if cryptographer  $i$  paid, then the protocol guarantees that the paying action is anonymous up to  $C \setminus \{j\}$  with respect to cryptographer  $j$ , as long as  $j \neq i$ .

**Coercion Resistance.** Voting protocols are protocols in which anonymity plays an important role. Voting protocols furthermore satisfy other interesting security properties. Aside from secrecy of votes (that every voter's choice should be private, and observers should not be able to figure out who voted how), other properties include fairness (voters do not have any knowledge of the distribution of votes until the final tallies are announced), verifiability (every voter should be able to check whether her vote was counted properly), and receipt freeness (no voter has the means to prove to another that she has voted in a particular manner).

This last property, receipt freeness, is particularly interesting in terms of epistemic content. Roughly speaking, receipt freeness says that a voter Alice cannot prove to a potential coercer Corinna that she voted in a particular way. This is the case even if Alice wishes to cooperate with Corinna; receipt freeness guarantees that such cooperation cannot lead to anything because it will be impossible for Corinna to be certain how Alice voted. In that sense, receipt freeness goes further than secrecy of votes. Even if Alice tells Corinna that she voted a certain way, Corinna has no way to verify Alice's assertion, and Alice has no way to convince her.

Coercion resistance is closely related to receipt freeness but is slightly stronger. Intuitively, a voting protocol is coercion resistant if it prevents voter coercion and vote buying even by active coercers: a coercer should not be able to influence the behavior of a voter.

Coercion resistance can be modeled epistemically, although the details of the modeling is subtle, and many important details will be skipped in the description below. Part of the difficulty and subtlety is that the idea of coercion means changing how a voter behaves based on a coercer’s desired outcome or goal, which needs to be modeled somehow.

One formalization of coercion resistance uses a model of voting protocols based on traces, where some specific agents are highlighted: a voter that the coercer tries to influence (called the coerced voter), the coercer, and the remaining agents and authorities, assumed to be honest. Every voter in the system votes according to a voting strategy, which in the case of honest voters is the strategy corresponding to the voting protocol.

The formalization assumes that every voter has a specific voting goal, formally captured by the set of traces in which that voter successfully votes according to her desired voting goal. The coercer, however, is intent on affecting the coerced voter—for instance, to coerce a vote for a given candidate, or perhaps to coerce a vote away from a given candidate. To coerce a voter, the coercer hands the coerced voter a particular strategy that will fulfill the coercer’s goals instead of the coerced voter’s. For instance, the coercer’s strategy may simply be one that forwards all messages to and from the coercer, effectively making the coerced voter a proxy for the coercer.

Let  $V$  be the space of possible strategies that voters and coercers can follow. Coercion resistance can be defined by saying that for every possible strategy  $v \in V$ , there is another strategy  $v' \in V$  that the coerced voter can use instead of  $v$  with the property that: (1) the voter always achieves her goal by using  $v'$ , and (2) the coercer does not know whether the coerced voter used strategy  $v$  or  $v'$ . In other words, in every trace in which the coerced voter uses strategy  $v$ , the coercer considers it possible, given her view of the trace, that the coerced voter is using strategy  $v'$  instead. Conversely, in every trace in which the coerced voter uses strategy  $v'$ , the coercer considers it possible that the coerced voter is using strategy  $v$ . So, the coercer cannot know whether the coerced voter followed the coercer’s instructions (i.e., used  $v$ ) or tried to achieve her own goal (i.e., used  $v'$ ). As in the case of information flow in event systems in §3.1, the definition of coercion resistance is a form of closure property on traces, which corresponds to lack of knowledge in the expected way, where knowledge is captured by an indistinguishability relation on states based on the coercer’s observations.

**Zero Knowledge.** The property *an agent does not learn anything about something*, as embodied in information-flow security policies and other forms of confidentiality, is generally modeled using an indistinguishability relation over states and enforced by making sure that there are enough states to prevent the confidential information from being known by unauthorized agents.

Another approach to modeling and enforcing this lack of learning is demonstrated by *zero knowledge interactive proof systems*. An interactive proof system for a string language  $L$  is a two-party system  $(P, V)$  in which a prover  $P$  tries to convince a verifier  $V$  that some string  $x$  is in  $L$  through a sequence of message exchanges amounting to an interactive proof of  $x \in L$ . Classically, the prover is assumed to be infinitely powerful, while the verifier is assumed to be a probabilistic polynomial-time Turing machine. An interactive proof system has the property that if  $x \in L$ , the conversation between  $P$  and  $V$  will show  $x \in L$  with high probability, and if  $x \notin L$ , the conversation between *any* prover and  $V$  will

show  $x \in L$  with low probability. (The details for why the second condition refers to any prover rather than just  $P$  is beyond the scope of this discussion.)

An interactive proof system for  $L$  is zero knowledge if whenever  $x \in L$  holds the verifier is able to generate *on its own* the conversations it would have had with the prover during an interactive proof of  $x \in L$ . The intuition here is that the verifier does not learn anything from a conversation with the prover (other than  $x \in L$ ) if it can learn exactly the same thing by generating that whole conversation itself. Thus, the only knowledge gained by the verifier is that which the prover initially set out to prove.

Zero knowledge interactive proof systems rely on indistinguishability, but not indistinguishability among a large set of states. Rather, it is indistinguishability between two scenarios: a scenario where the verifier interacts with the prover, and a scenario where the verifier does not interact with the prover but instead simulates a complete interaction with the prover. This simulation paradigm, a core notion in modern theoretical computer science, says roughly that an agent does not gain any knowledge from interacting with the outside world if she can achieve the same results without interacting with the outside world.

To give a sense of the kinds of definitions that arise in this context, here is one formal definition of perfect zero knowledge:

Let  $(P, V)$  be an interactive proof system for  $L$ , where  $P$  (the prover) is an interactive Turing machine and  $V$  (the verifier) is a probabilistic polynomial-time interactive Turing machine. System  $(P, V)$  is *perfect zero-knowledge* if for every probabilistic polynomial-time interactive Turing machine  $V^*$  there is a probabilistic polynomial-time Turing machine  $M^*$  (the simulator) such that for every  $x \in L$  the following two random variables are identically distributed:

- (i) the output of  $V^*$  interacting with  $P$  on common input  $x$ ;
- (ii) the output of machine  $M^*$  on input  $x$ .

While the details are beyond the scope of this chapter, the intuition behind this definition is to have, for every possible verifier  $V^*$  (and not only  $V$ ) interacting with  $P$ , a machine  $M^*$  that can simulate the interaction of  $V^*$  and  $P$  even though it does not have access to the prover  $P$ . The existence of such simulators implies that  $V^*$  does not gain any knowledge from  $P$ .

This gives a different epistemic foundation for confidentiality, one that is intimately tied to computation and its complexity. The relationship with classical epistemic logic is essentially unexplored.

## 5 Perspectives

The preceding sections illustrate how extensively epistemic concepts, explicitly framed as an epistemic logic or not, have been applied to security research. Whether the application of these concepts has been successful is a more subjective question.

In a certain sense, the problems described in this chapter are solved problems by now. Confidentiality and authentication in cryptographic protocol analysis under a formal model

of cryptography and Dolev-Yao attackers, for example, can be checked quite efficiently with a vast array of methods, at least for common security properties, and the definitions used approximate the epistemic definitions quite closely.

So what are the remaining challenges in cryptographic protocol analysis, and has epistemic logic a role to play? The most challenging aspect of cryptographic protocol analysis is to move beyond Dolev-Yao attackers and beyond formal models of cryptography, towards more concrete models of cryptography.

Moving beyond a Dolev-Yao attacker requires shifting the notion of message knowledge to use richer algebras of message with more operations. Directions that have been explored include providing attackers with the ability to perform offline dictionary attacks, working with an XOR operation, or even number-based operations such as exponentiation. One problem is that when the algebra of messages is subject to too many algebraic properties, determining whether an attacker knows a message may quickly become undecidable. Even when message knowledge for an attacker is decidable, it may still be too complex for efficient reasoning. It is not entirely clear how epistemic concepts can help solve problems in that arena.

Moving from a formal model of cryptography to a concrete model, one that reflects real encryption schemes more accurately using sequences of bits and computational indistinguishability, requires completely shifting the approach to cryptographic protocol analysis.

Formal models of cryptography work by abstracting away the *one-way security* property of encryption schemes—that it is computationally hard to recover the sourcetext from a ciphertext without knowing the encryption key. More concrete models of cryptography rely on stronger properties than one-way security, properties such as *semantic security*, which intuitively says that if any information about a message  $m$  can be computed by an efficient algorithm given the ciphertext  $e_k(m)$  for a random  $k$  and  $m$  chosen according to an arbitrary probability distribution, that same information can be computed without knowing the ciphertext. In other words, the ciphertext  $e_k(m)$  offers no advantage in computing information about some message  $m$  chosen from an arbitrary probability distribution.

The definition of semantic security is reminiscent of the definition of zero knowledge interactive proof systems in §4, and it is no accident, as they both rely on a simulation paradigm to express the fact that no knowledge is gained. As in the case of zero knowledge interactive proof systems, there is a clear epistemic component to the definition of semantic security, one to which classical epistemic logic has not been applied.

The main difficulty with applying classical epistemic logic to concrete models of cryptography is that these models take attackers to be probabilistic polynomial-time Turing machines, and take security properties to be probabilistic properties relative to those probabilistic polynomial-time Turing machines. This means that an epistemic approach to concrete models of cryptography needs to be probabilistic as well as computationally bounded. The former is not a problem, since probabilistic reasoning shares much of the same foundations as epistemic reasoning. But the latter is more complicated. Concrete models of cryptography are not based on impossibility, but on computational hardness. And while possible-worlds definitions of knowledge are well suited to talking about impossible versus possible outcomes, they fare less well at talking about difficult versus easy outcomes.

The trouble that possible-worlds definitions of knowledge run into when trying to in-

corporate a notion of computational difficulty is really the problem of logical omniscience in epistemic logic under a different guise. Agents, under standard possible-worlds definitions of knowledge, know all tautologies, and know all logical consequences of their knowledge: if  $K\varphi$  is true and  $\varphi \Rightarrow \psi$  is valid, then  $K\psi$  is also true. Any normal epistemic operator will satisfy these properties, and in particular, any epistemic logic based on Kripke structures will satisfy these properties. Normality does not deal well with computational difficulty, because while it may be computationally difficult to establish that  $\varphi \Rightarrow \psi$  is valid, a normal modal logic will happily derive all knowledge-based consequences of that valid formula. It would seem that giving a satisfactory epistemic account of concrete models of cryptography requires a non-normal epistemic logic, one that supports a form of resource-bounded knowledge. Resource-bounded knowledge is not well understood, and logics for resource-bounded knowledge still feel too immature to form a solid basis for reasoning about concrete models of cryptography.

Leaving aside concrete models of cryptography, it is almost impossible to discuss epistemic logic in the context of cryptographic protocols without addressing the issue of BAN Logic. BAN Logic is an interesting and original use of logic, developed to prove cryptographic protocol properties manually by paralleling informal arguments for protocol correctness.

BAN Logic has spilled a lot of virtual ink. Aside from its technical limitations—it requires a protocol idealization step that remains outside the purview of the logic but affects the results of analysis—the logic is considered somewhat *passé*. Other approaches we saw in §2.5 operate in the same space, namely analyzing cryptographic protocols under a formal model of cryptography in the presence of Dolev-Yao attackers, and most are less limited and more easily automated. Other approaches, such as Protocol Composition Logic, even advocate Hoare-style reasoning about the protocol text from within the logic, just like BAN Logic.

My perspective on BAN Logic is that it tried to do something which has not really been tried since, something that remains a sort of litmus test for our understanding of security in cryptographic protocols: identifying high-level primitives that capture relevant concepts for security, high-level primitives that match our intuitive understanding of security properties, those same intuitions that guide our design of cryptographic protocols in the first place. We do not have such high-level primitives in any other framework, all of which tend to work at much lower levels of abstraction. The primitives in BAN Logic are intuitively attractive, but poorly understood. The continuing conversation on BAN Logic is a reminder that we still do not completely understand the basic concepts and basic terms needed to discuss cryptographic protocols, and I think BAN Logic remains relevant, if only as a nagging voice telling us that we have not quite gotten it right yet.

Many of the issues that arise when trying to push cryptographic protocol analysis from a formal model of cryptography to a more concrete model also come up in the context of information-flow security. As mentioned in §3.2, recent work has turned to the question of declassification, or controlled release of information. The reason for this is purely pragmatic: most applications need to release some kind of information in order to do any useful work, even under a lax interpretation of noninterference.

But it does not take long to see that even a controlled release of information can lead to unwanted release of information in the aggregate. Returning to the password-login

problem from §3.2, it is clear that every wrong attempt at entering a password leaks some information, something that needs to happen if the login screen is to operate properly. But of course, repeated attempts at checking the password will eventually lead to the correct password as the only remaining possibility, which is a severe undesirable release of information. Security policies controlling declassification therefore seem to require a way to account for more quantitative notions of leakage which aggregate over time, something that symbolic approaches to information-flow security have difficulty handling well.

Modeling information flow quantitatively can be seen as a move from reasoning about information as a monolithic unit to reasoning about information as a resource. Once we make that leap, other resources affecting information flow start suggesting themselves. For example, execution time can leak information. Consider the simple program:

```

if (high-security Boolean variable)
  then fast code
  else slow code
```

By observing the execution time of the program, we can determine the value of the high-security Boolean variable. This example is rather silly, of course, but it illustrates the point that information leakage can occur based on observations of other resources than simply the state of memory.

What about the combination of information flow and cryptography? After all, in practice, systems do use cryptography internally to help keep data confidential. Encrypted data can presumably be written on shared storage (which might be easier to manage than storage segregated into high-security and low-security storage) or moved online, or in general given to low-security agents without information being released, as long as they do not have the key or the resources to decrypt. Accounting for cryptography in information-flow security raises questions similar to those in cryptographic protocol analysis concerning what models of cryptography to use and how to account for the cryptographic capabilities of attackers. It also raises difficulties similar to those in cryptographic protocol analysis when trying to move from a formal model of cryptography to a concrete model, including how to provide an epistemic foundation for information flow using a resource-bounded definition of knowledge.

**Conclusion.** Epistemic concepts are central to many aspects of reasoning about security. In some cases, these epistemic concepts may even naturally take expression in a *bona fide* epistemic logic that can be used to formalize the reasoning. But whether an epistemic logic is used or not, the underlying concepts are clearly epistemic. In particular, the notion of truth at all possible worlds reappears in many different guises throughout the literature.

Research in security analysis has reached a sort of convergence point around the use of symbolic methods. The challenge seems to be to move beyond this convergence point, and such a move requires taking resources seriously: realistic definitions of security rely on the notion that exploiting a vulnerability should require more resources (time, power, information) than are realistically available to an attacker. In epistemic terms, what is needed is a reasonably well-behaved definition of resource-bounded knowledge, itself an

active area of research in epistemic logic. It would appear, then, that advances in epistemic logic may well help increase our ability to reason about security in direct ways.

**Acknowledgments.** Thanks to Aslan Askarov, Philippe Balbiani, Stephen Chong, and Vicky Weissman for comments on an early draft of this chapter.

## 6 Bibliographic Notes and Further Reading

For the basics of epistemic logic, both the syntax and the semantics, the reader is referred to the introductory chapter of the current volume. For the sake of making this chapter as self-contained as possible, most of the background material can be usefully obtained from the textbooks of Fagin, Halpern, Moses, and Vardi [42] and Meyer and Van der Hoek [88]. The possible-worlds definition of knowledge used throughout this chapter is simply the view that knowledge is truth at all worlds that an agent considers as possible alternatives to the current world, a view which goes back to Hintikka [64].

**Cryptographic Protocols.** While the focus of the section is on symbolic cryptographic protocol analysis, cryptographic protocols can also be studied from the perspective of more computationally-driven cryptography, of the kind described in §5; see Goldreich [49]. The Russian Cards problem, which was first presented at the Moscow Mathematic Olympiads in 2000, is described formally and studied from an epistemic perspective by Van Ditmarsch [122]. The problem has been used as a benchmark for several epistemic logic model checkers [123]. The Dining Cryptographers problem and the corresponding protocol is described by Chaum [27]. It was proved correct in an epistemic temporal logic model checker by Van der Meyden and Su [121].

For a good overview of classical cryptography along with some perspectives on protocols, see Stinson [110] and Schneier [107]; both volumes contain descriptions of DES, AES, RSA, and elliptic-curve cryptography. Goldreich [48, 49] is also introductory, but from the perspective of modern computational cryptography.

For a good high-level survey of the kind of problems surrounding the design and deployment of cryptographic protocols, see Anderson and Needham [8], then follow up with Abadi and Needham's [5] prudent engineering practices. The key distribution protocol used as the first example in §2.2 is related to the Yahalom protocol described by Burrows, Abadi, and Needham [23]. The Needham-Schroeder protocol was first described in Needham and Schroeder [94]. The man-in-the-middle attack on the Needham-Schroeder protocol in the presence of an insider attacker was pointed out by Lowe [77], and the fix was analyzed by Lowe [78].

The Dolev-Yao model of the attacker given in §2.3 is due to Dolev and Yao [38].

The formal definition of message knowledge via *Analyzed* and *Synthesized* sets is taken from Paulson [98]. Equivalent definitions are given in nearly every formal system for reasoning about cryptographic protocols in a formal model of cryptography. Message knowledge can be defined using a local deductive system, which makes it fit nicely within the deductive knowledge framework of Konolige [71]—see also Pucella [99]. More generally,

message knowledge is a form of algorithmic knowledge [56], that is, a local form of knowledge that relies on an algorithm to compute what an agent knows based on the local state of the agent. In the case of a Dolev-Yao attacker, this local algorithm simply computes the sets of analyzed and synthesized messages [61].

Another way of defining message knowledge is the hidden automorphism model, due to Merritt [87], which is a form of possible-worlds knowledge. While it never gained much traction, it has been used in later work by Toussaint and Wolper [119] and also in the logic of Bieber [20]. It uses algebraic presentations of encryption schemes called *cryptoalgebras*. There is a unique surjective cryptoalgebra homomorphism from the free cryptoalgebra over a set of plaintexts and keys to any cryptoalgebra over the same plaintexts and keys which acts as the identity on plaintexts and keys. Message knowledge in a given cryptoalgebra  $C$  is knowledge of the structure of messages as given by that surjective homomorphism from the free cryptoalgebra to  $C$ . A revealed reduct is a subset of  $C$  that the agent has seen. A state of knowledge with respect to revealed reduct  $R$  is a set of mappings  $f$  from the free cryptoalgebra to  $C$  that are homomorphisms on  $f^{-1}(R)$ . In this context, an agent knows message  $m$  if the agent knows the representation of message  $m$ , meaning that  $m$  is the image of the same free cryptoalgebra term under every mapping in the state of knowledge of the agent. Thus, if an agent receives  $\{m_1\}_k$  and  $\{m_2\}_k$  but does not receive  $k$ , then only  $\{m_1\}_k$  and  $\{m_2\}_k$  are in the revealed reduct; the agent may consider any distinct messages  $m'_1$  and  $m'_2$  to map to  $\{m_1\}_k$  and  $\{m_2\}_k$  after encryption with  $k$ , since any such mapping will act as a homomorphism on the pre-image of the revealed reduct.

Possible-worlds definitions of knowledge in the presence of cryptography are problematic because cryptography affects what agents can observe, and this impacts the definition of the accessibility relation between worlds. The idea of replacing encrypted messages in the local state of agents by a token goes back to Abadi and Tuttle's semantics for BAN Logic [7]. Treating encrypted messages as tokens while still allowing agents to distinguish different encrypted messages is less common, but has been used at least by Askarov and Sabelfeld [10] and Askarov, Hedin, and Sabelfeld [9] in the context of information flow.

There are several frameworks for formally reasoning about cryptographic protocols, and I shall not list them all here. But I hope to provide enough pointers to the literature to ensure that the important ones are covered. For an early survey on the state of the art in formal reasoning about cryptographic protocols until 1995, see Meadows [85].

The Inductive Method described in §2.5 is due to Paulson [98], and is built atop the Isabelle logical framework [97], a framework for higher-order logic. BAN Logic is introduced by Burrows, Abadi, and Needham [23], who use it to perform an analysis of several existing protocols in the literature. The logic courted controversy pretty much right from the start [95, 24]. Probably the most talked-about problem with BAN Logic is the lack of an independently-motivated semantics which would ensure that statements of the logic match operational intuition. Without such a semantics, it is difficult to argue for the reasonableness of the result of a BAN Logic analysis, except for the pragmatic observation that failure to prove a statement in BAN Logic often indicates a problem with the cryptographic protocol. Abadi and Tuttle [7] attempt to remedy the situation by defining a semantics for BAN Logic. Successor logics extending or modifying BAN Logic generally start with a variant of the Abadi-Tuttle semantics [113, 51, 125, 116, 126, 111]. Contemporary epistemic logic alternatives to BAN Logic were also developed, using a semantics in



terms of protocol execution, but they never really took hold [20, 92].

The model checker MCK is described by Gammie and Van der Meyden [43], and was used to analyze the Dining Cryptographers protocol [121] as well as the Seven Hands protocol for the Russian Cards problem [123]. TDL is an alternative epistemic temporal logic for reasoning about cryptographic protocols with a model checker developed by Penczek and Lomuscio [76], based on a earlier model checker [100]. TDL is a branching-time temporal epistemic logic extended with a message knowledge primitive in addition to standard possible-worlds knowledge for expressing higher-order knowledge, and does not provide explicit support for attackers in its modeling language. The model-checking complexity results mentioned are due to Van der Meyden and Shilov [120]; see also Engelhardt, Gammie, and Van der Meyden [40] and Huang and Van der Meyden [68].

Another epistemic logic which forms the basis for reasoning about cryptographic protocol is Dynamic Epistemic Logic (DEL) [45]. DEL is an epistemic logic of broadcast announcements which includes formulas of the form  $[\rho]_i \varphi$ , read  *$\varphi$  holds after agent  $i$  broadcasts formula  $\rho$* , where  $\rho$  is a formula in a propositional epistemic sublanguage. (The actual syntax of DEL is slightly different.) Agents may broadcast that they know a fact, and this broadcast affects the knowledge of other agents. Kripke structures are used to capture the state of knowledge of agents at a point in time, and agent  $i$  announcing  $\rho$  will change Kripke structure  $M$  representing the current state of knowledge of all agents into a Kripke structure  $M^{\rho,i}$  representing the new state of knowledge that obtains. Dynamic Epistemic Logic has been used to analyze the Seven Hands protocol in great detail [122]. Extensions to handle cryptography are described by Hommersom, Meyer, and De Vink [66], as well as Van Eijck and Orzan [124].

Process calculi, starting with the spi calculus [4] and later the applied pi calculus [3], have been particularly successful tools for reasoning about cryptographic protocols. These use either observational equivalence to show that a process implementation of the protocol is equivalent to another process that clearly satisfies the required properties, or static analysis such as type checking to check the properties [52]. Epistemic logics defined against models obtained from processes are given by Chadha, Delaune, and Kremer [26] and Toninho and Caires [118]. Another process calculus, CSP, has also proved popular as a foundation for reasoning about cryptographic protocols [79, 102].

Finally, other approaches rely on logic programming ideas: the NRL protocol analyzer [86], Multiset Rewriting [25], and ProVerif [21]. Thayer, Herzog, and Guttman [117] introduce a distinct semantic model for protocols, strand spaces, which has some advantages over traces. Syverson [114] develops an authentication logic on top of strand spaces, while Halpern and Pucella [59] investigate the suitability of strand spaces as a basis for epistemic reasoning.

**Information Flow.** Bell and LaPadula [18, 73] were among the first to develop mandatory access control, and introducing the idea of attaching security levels to data to enforce confidentiality.

Early work on information flow security mostly focused on event traces, and tried to describe both closure conditions on traces, as well as unwinding conditions that would allow one to check that a set of event traces satisfies the security condition. Separability was defined by McLean [84], noninference by O'Halloran [96], generalized noninference

by McLean [84], and generalized noninterference by McCullough [82, 83] following the work of Goguen and Meseguer [46, 47]. Other definitions of information-flow security for event systems are given by Sutherland [112] and Wittbold and Johnson [127]. A modern approach to information-flow security in event systems is described by Mantel [81]. The set-theoretic description of the knowledge operator is taken from Halpern [54], but appears in various guises in the economics literature [11]. Halpern and O’Neill [58] layer an explicit epistemic language on top of the event models re-expressed as Kripke structures, and show that the resulting logic can capture common definitions of confidentiality in event systems.

Denning and Denning [36] first pointed out that programming languages are a useful setting for reasoning about information flow by observing that static analysis can be used to identify and control information flow. Most recent work on information-flow security from a programming language perspective goes back to Heintze and Riecke’s Secure Lambda Calculus [63] in a functional language setting, and Smith and Volpano [109] in an imperative language setting. Honda, Vasconcelos, and Yoshida [67] give a similar development in the context of a process calculus. Sabelfeld and Myers [104] give a survey and overview of the state of the field up to 2003. Balliu, Dam, and Le Guernic [14, 15] offer a rare use of an explicit epistemic temporal logic to reason about information-flow security. Sabelfeld and Sands [105] give a good overview of the issues involved in declassification for language-based information flow. Askarov and Sabelfeld [10] use an epistemic logic in the context of declassification. Chong [29] uses a form of algorithmic knowledge to model information release requirements.

**Beyond Confidentiality.** Protocols for anonymous communication generally rely on a cloud of intermediaries that prevent information about the identity of the original sender to be isolated; Crowds is an example of such a protocol [101]. Anonymity has been well studied as an instance of confidentiality [69, 44]. The explicit connection with epistemic logic was made by Halpern and O’Neill [57], which is the source of the definitions in §4. An early analysis of anonymity via epistemic logic is given by Syverson and Stubblebine [115].

Anonymity is an important component of voting protocols. Van Eijck and Orzan [124] prove anonymity for a specific voting protocol using epistemic logic. More general analyses of voting protocols with epistemic logic have also been attempted [16, 72]. The model of coercion resistance in §4 is from Küsters and Truderung [72].

Zero knowledge interactive proof systems were introduced by Goldwasser, Micali, and Rackoff [50] and have become an important tool in theoretical computer science. A good overview is given by Goldreich [48]. Halpern, Moses, and Tuttle [55] give an epistemically-motivated analysis of zero knowledge interactive proof systems using a computationally-bounded definition of knowledge devised by Moses [93].

Another context in which epistemic concepts—or perhaps more accurately, epistemic vocabulary—appear is that of authorization and trust management. Credential-based authorization policies can be used to control access to resources by requiring agents to present appropriate credentials (such as certificates) proving that they are allowed access. Because systems that rely on credential-based authorization policies are often decentralized systems, meaning that there is no central clearinghouse for determining for every authorization request whether an agent has the appropriate credentials, the entire approach relies

on a web of trust between agents and credentials. Since in many such systems credentials can be delegated—an agent may allow another agent to act on her behalf—not only can credential checking become complicated, but authorization policies themselves become nontrivial to analyze to determine contradictions (an action being both allowed and forbidden by the policy under certain conditions) or coverage (a class of actions remaining unregulated by the policy under certain conditions). Where do epistemic concepts come up in such a scenario? Authorization logics from the one introduced by Abadi, Burrows, Lampson, and Plotkin [1] to the recent NAL [106] have been described as logics of belief, and are somewhat reminiscent of BAN Logic. One of their basic primitives is a formula  $A \text{ says } F$ , which as a credential means that  $A$  believes and is accountable for the truth of  $F$ . Delegation, for example, is captured by a formula  $(A \text{ says } F) \Rightarrow (B \text{ says } G)$ . This form of belief is entirely axiomatic, just like belief in BAN Logic.

**Perspectives.** Ryan and Schneider [103] have extended the Dolev-Yao model of attackers with an XOR operation; Millen and Shmatikov [90] with products and enough exponentiation to model the Diffie-Hellman key-establishment protocol [37]; and Lowe [80] and later Corin, Doumen, and Etalle [32] and Baudet [17] with the ability to mount offline dictionary attacks. As described by Halpern and Pucella [61], many of these can be expressed using algorithmic knowledge, at least in the context of eavesdropping attackers. More generally, extending Dolev-Yao with additional operations can best be studied using equational theories, that is, equations induced by looking at the algebra of the additional operations; see for example Abadi and Cortier [2] and Chevalier and Rusinowitch [28].

While it would be distracting to discuss the back and forth over BAN Logic in the decades since its inception, I will point out that recent work by Cohen and Dam has taken a serious look at the logic with modern eyes, and highlighted both interesting interpretations as well as subtleties [31, 30]. The protocol composition logic PCL of Datta, Derek, Mitchell, and Roy [33], which builds on earlier work by Durgin, Mitchell, and Pavlovic [39], is a modern attempt at devising a logic for Hoare-style reasoning about cryptographic protocols.

A good overview of concrete models of cryptography is given by Goldreich [48]. Semantic security, among others, is studied by Bellare, Chor, Goldreich, and Schnorr [19]. The relationship between formal models of cryptography and concrete models—how well does the former approximate the latter?—has been explored by Abadi and Rogaway [6], and later extended by Micciancio and Warinschi [89], among others. Backes, Hofheinz, and Unruh [12] provide a good overview.

Approaches to analyze cryptographic protocols in a concrete model of cryptography have been developed [75, 91]. In recent years some of the approaches for analyzing cryptographic protocols in a formal model of cryptography have been modified to work with a concrete model of cryptography, such as PCL [34] and ProVerif [22]. In some cases, cryptographic protocol analysis in a concrete model relies on extending indistinguishability over states to indistinguishability over the whole protocol [35].

Defining a notion of resource-bounded knowledge that does not suffer from the logical omniscience problem is an ongoing research project in the epistemic logic community, and various approaches have been advocated, each with its advantages and its deficiencies: algorithmic knowledge [56], impossible possible worlds [65], awareness [41]. A comparison

between the approaches in terms of expressiveness and pragmatics appears in Halpern and Pucella [60].

Information flow in probabilistic programs was first investigated by Gray and Syver-son [53] using probabilistic multiagent systems [62]. Backes and Pfitzmann [13] study it in a more computational setting. Smith [108] presents some of the tools that need to be considered to analyze the kind of partial information leakage occurring in the password-checking example. Preliminary work on information flow in the presence of cryptography includes Laud [74], Hutter and Schairer [70], and Askarov, Hedin, and Sabelfeld [9].

## References

- [1] M. Abadi, M. Burrows, B. Lampson, and G. Plotkin. A calculus for access control in distributed systems. *ACM Transactions on Programming Languages and Systems*, 15(4):706–734, 1993.
- [2] M. Abadi and V. Cortier. Deciding knowledge in security protocols under equational theories. In *Proc. 31st Colloquium on Automata, Languages, and Programming (ICALP’04)*, volume 3142 of *Lecture Notes in Computer Science*, 2004.
- [3] M. Abadi and C. Fournet. Mobile values, new names, and secure communication. In *Proc. 28th Annual ACM Symposium on Principles of Programming Languages (POPL’01)*, pages 104–115, 2001.
- [4] M. Abadi and A. D. Gordon. A calculus for cryptographic protocols: The spi calculus. *Information and Computation*, 148(1):1–70, 1999.
- [5] M. Abadi and R. Needham. Prudent engineering practice for cryptographic protocols. *IEEE Transactions on Software Engineering*, 22(1):6–15, 1996.
- [6] M. Abadi and P. Rogaway. Reconciling two views of cryptography (the computational soundness of formal encryption). *Journal of Cryptology*, 15(2):103–127, 2002.
- [7] M. Abadi and M. R. Tuttle. A semantics for a logic of authentication. In *Proc. 10th ACM Symposium on Principles of Distributed Computing (PODC’91)*, pages 201–216, 1991.
- [8] R. Anderson and R. Needham. Programming Satan’s computer. In J. van Leeuwen, editor, *Computer Science Today: Recent Trends and Developments*, volume 1000 of *Lecture Notes in Computer Science*, pages 426–440. Springer-Verlag, 1995.
- [9] A. Askarov, D. Hedin, and A. Sabelfeld. Cryptographically-masked flows. *Theoretical Computer Science*, 402(2–3):82–101, 2008.
- [10] A. Askarov and A. Sabelfeld. Gradual release: Unifying declassification, encryption and key release policies. In *Proc. 2007 IEEE Symposium on Security and Privacy*, pages 207–221. IEEE Computer Society Press, 2007.
- [11] R. J. Aumann. Notes on interactive epistemology. Cowles Foundation for Research in Economics working paper, 1989.
- [12] M. Backes, D. Hofheinz, and D. Unruh. CoSP: A general framework for computational soundness proofs. In *Proc. 16th ACM Conference on Computer and Communications Security (CCS’09)*, pages 66–78. ACM Press, 2009.
- [13] M. Backes and B. Pfitzmann. Computational probabilistic non-interference. In *Proc. 7th European Symposium on Research in Computer Security (ESORICS’02)*, volume 2502 of *Lecture Notes in Computer Science*, pages 1–23. Springer-Verlag, 2002.

- [14] M. Balliu, M. Dam, and G. Le Guernic. Epistemic temporal logic for information flow security. In *Proc. ACM SIGPLAN 6th Workshop on Programming Languages and Analysis for Security (PLAS'11)*. ACM Press, 2011.
- [15] M. Balliu, M. Dam, and G. Le Guernic. ENCOVER: Symbolic exploration for information flow security. In *Proc. 25th IEEE Computer Security Foundations Symposium (CSF'12)*, pages 30–44. IEEE Computer Society Press, 2012.
- [16] A. Baskar, R. Ramanujam, and S. P. Suresh. Knowledge-based modelling of voting protocols. In *Proc. 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'07)*, pages 62–71. ACM Press, 2007.
- [17] M. Baudet. Deciding security of protocols against off-line guessing attacks. In *Proc. 12th ACM Conference on Computer and Communications Security (CCS'05)*, pages 16–25. ACM Press, 2005.
- [18] D. E. Bell and L. J. LaPadula. Secure computer systems: Mathematical foundations. Technical Report MTR-2547, Volume 1, MITRE Corporation, 1973.
- [19] M. Bellare, B. Chor, O. Goldreich, and C. Schnorr. Relations among notions of security for public-key encryption schemes. In *Proc. 18th Annual International Cryptology Conference (CRYPTO'98)*, volume 1462 of *Lecture Notes in Computer Science*, pages 26–45. Springer-Verlag, 1998.
- [20] P. Bieber. A logic of communication in hostile environment. In *Proc. 3rd IEEE Computer Security Foundations Workshop (CSFW'90)*, pages 14–22. IEEE Computer Society Press, 1990.
- [21] B. Blanchet. An efficient cryptographic protocol verifier based on Prolog rules. In *Proc. 14th IEEE Computer Security Foundations Workshop (CSFW'01)*, pages 82–96. IEEE Computer Society Press, 2001.
- [22] B. Blanchet. A computationally sound mechanized prover for security protocols. *IEEE Transactions on Dependable and Secure Computing*, 5(4):193–207, 2008.
- [23] M. Burrows, M. Abadi, and R. Needham. A logic of authentication. *ACM Transactions on Computer Systems*, 8(1):18–36, 1990.
- [24] M. Burrows, M. Abadi, and R. Needham. Rejoinder to Nessett. *ACM Operating Systems Review*, 24(2):39–40, 1990.
- [25] I. Cervesato, N. Durgin, P. Lincoln, J. Mitchell, and A. Scedrov. A meta-notation for protocol analysis. In *Proc. 12th IEEE Computer Security Foundations Workshop (CSFW'99)*, pages 55–69. IEEE Computer Society Press, 1999.
- [26] R. Chadha, S. Delaune, and S. Kremer. Epistemic logic for the applied pi calculus. In *Proc. IFIP International Conference on Formal Techniques for Distributed Systems*, volume 5522 of *Lecture Notes in Computer Science*, pages 182–197. Springer-Verlag, 2009.
- [27] D. Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1(1):65–75, 1988.
- [28] Y. Chevalier and M. Rusinowitch. Hierarchical combination of intruder theories. *Information and Computation*, 206(2–4):352–377, 2008.
- [29] S. Chong. Required information release. In *Proc. 23rd IEEE Computer Security Foundations Symposium (CSF'10)*, pages 215–227. IEEE Computer Society Press, 2010.
- [30] M. Cohen and M. Dam. A completeness result for BAN logics. In *Proc. Methods for Modalities (M4M)*, pages 202–219, 2005.

- [31] M. Cohen and M. Dam. Logical omniscience in the semantics of BAN logic. In *Proc. Workshop on Foundations of Computer Security (FCS'05)*, pages 121–132, 2005.
- [32] R. Corin, J. Doumen, and S. Etalle. Analysing password protocol security against off-line dictionary attacks. In *Proc. 2nd International Workshop on Security Issues with Petri Nets and other Computational Models (WISP'04)*, volume 121 of *Electronic Notes in Theoretical Computer Science*, pages 47–63. Elsevier Science Publishers, 2005.
- [33] A. Datta, A. Derek, J. C. Mitchell, and A. Roy. Protocol Composition Logic (PCL). *Electronic Notes in Theoretical Computer Science*, 172(1):311–358, 2007.
- [34] A. Datta, A. Derek, J. C. Mitchell, V. Shmatikov, and M. Turuani. Probabilistic polynomial-time semantics for a protocol security logic. In *Proc. 32nd Colloquium on Automata, Languages, and Programming (ICALP'05)*, pages 16–29, 2005.
- [35] A. Datta, R. Küsters, J. C. Mitchell, A. Ramanathan, and V. Shmatikov. Unifying equivalence-based definitions of protocol security. In *Proc. Workshop on Issues in the Theory of Security (WITS'04)*, 2004.
- [36] D. E. Denning and P. J. Denning. Certification of programs for secure information flow. *Communications of the ACM*, 20(7):504–513, 1977.
- [37] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22:664–654, 1976.
- [38] D. Dolev and A. C. Yao. On the security of public key protocols. *IEEE Transactions on Information Theory*, 29(2):198–208, 1983.
- [39] N. A. Durgin, J. C. Mitchell, and D. Pavlovic. A compositional logic for proving security properties of protocols. *Journal of Computer Security*, 11(4):677–722, 2003.
- [40] K. Engelhardt, P. Gammie, and R. van der Meyden. Model checking knowledge and linear time: PSPACE cases. In *Proc. Symposium on Logical Foundations of Computer Science*, volume 4514 of *Lecture Notes in Computer Science*, pages 195–211. Springer-Verlag, 2007.
- [41] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.
- [42] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
- [43] P. Gammie and R. van der Meyden. MCK: Model checking the logic of knowledge. In *Proc. 16th International Conference on Computer Aided Verification (CAV'04)*, *Lecture Notes in Computer Science*, pages 479–483. Springer-Verlag, 2004.
- [44] F. D. Garcia, I. Hasuo, W. Pieters, and P. van Rossum. Provable anonymity. In *Proc. 3rd ACM Workshop on Formal Methods in Security Engineering (FMSE 2005)*, pages 63–72. ACM Press, 2005.
- [45] J. Gerbrandy. Dynamic epistemic logic. In *Logic, Language, and Information*, volume 2. CSLI Publication, 1999.
- [46] J. A. Goguen and J. Meseguer. Security policies and security models. In *Proc. 1982 IEEE Symposium on Security and Privacy*, pages 11–20. IEEE Computer Society Press, 1982.
- [47] J. A. Goguen and J. Meseguer. Unwinding and inference control. In *Proc. 1984 IEEE Symposium on Security and Privacy*, pages 75–86. IEEE Computer Society Press, 1984.
- [48] O. Goldreich. *Foundations of Cryptography: Volume 1, Basic Tools*. Cambridge University Press, 2001.

- [49] O. Goldreich. *Foundations of Cryptography: Volume 2, Basic Applications*. Cambridge University Press, 2004.
- [50] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989.
- [51] L. Gong, R. Needham, and R. Yahalom. Reasoning about belief in cryptographic protocols. In *Proc. 1990 IEEE Symposium on Security and Privacy*, pages 234–248. IEEE Computer Society Press, 1990.
- [52] A. D. Gordon and A. Jeffrey. Authenticity by typing for security protocols. *Journal of Computer Security*, 11(4):451–520, 2003.
- [53] J. W. Gray, III and P. F. Syverson. A logical approach to multilevel security of probabilistic systems. *Distributed Computing*, 11(2):73–90, 1998.
- [54] J. Y. Halpern. Set-theoretic completeness for epistemic and conditional logic. *Annals of Mathematics and Artificial Intelligence*, 26:1–27, 1999.
- [55] J. Y. Halpern, Y. Moses, and M. R. Tuttle. A knowledge-based analysis of zero knowledge. In *Proc. 20th Annual ACM Symposium on the Theory of Computing (STOC’88)*, pages 132–147, 1988.
- [56] J. Y. Halpern, Y. Moses, and M. Y. Vardi. Algorithmic knowledge. In *Proc. 5th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK’94)*, pages 255–266. Morgan Kaufmann, 1994.
- [57] J. Y. Halpern and K. O’Neill. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–514, 2005.
- [58] J. Y. Halpern and K. O’Neill. Secrecy in multiagent systems. *ACM Transactions on Information and System Security*, 12(1), 2008.
- [59] J. Y. Halpern and R. Pucella. On the relationship between strand spaces and multi-agent systems. *ACM Transactions on Information and System Security*, 6(1):43–70, 2003.
- [60] J. Y. Halpern and R. Pucella. Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial Intelligence*, 175(1):220–235, 2011.
- [61] J. Y. Halpern and R. Pucella. Modeling adversaries in a logic for security protocol analysis. *Logical Methods in Computer Science*, 8(1:21), 2012.
- [62] J. Y. Halpern and M. R. Tuttle. Knowledge, probability, and adversaries. *Journal of the ACM*, 40(4):917–962, 1993.
- [63] N. Heintze and J. G. Riecke. The SLam calculus: Programming with secrecy and integrity. In *Proc. 25th Annual ACM Symposium on Principles of Programming Languages (POPL’98)*, pages 365–377. ACM Press, 1998.
- [64] J. Hintikka. *Knowledge and Belief*. Cornell University Press, 1962.
- [65] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.
- [66] A. Hommersom, J.-J. Meyer, and E. de Vink. Update semantics of security protocols. *Synthese*, 142(2):229–267, 2004.
- [67] K. Honda, V. Vasconcelos, and N. Yoshida. Secure information flow as typed process behaviour. In *European Symposium on Programming*, volume 1782 of *Lecture Notes in Computer Science*, pages 180–199. Springer-Verlag, 2000.

- [68] X. Huang and R. van der Meyden. The complexity of epistemic model checking: Clock semantics and branching time. In *Proc. 19th European Conference on Artificial Intelligence (ECAI'10)*, pages 549–554. IOS Press, 2010.
- [69] D. Hughes and V. Shmatikov. Information hiding, anonymity and privacy: A modular approach. *Journal of Computer Security*, 12(1):3–36, 2004.
- [70] D. Hutter and A. Schairer. Possibilistic information flow control in the presence of encrypted communication. In *Proc. 9th European Symposium on Research in Computer Security (ESORICS'04)*, volume 3193 of *Lecture Notes in Computer Science*, pages 209–224. Springer-Verlag, 2004.
- [71] K. Konolige. *A Deduction Model of Belief*. Morgan Kaufmann, 1986.
- [72] R. Küsters and T. Truderung. An epistemic approach to coercion-resistance for electronic voting protocols. In *Proc. 2009 IEEE Symposium on Security and Privacy*, pages 251–266. IEEE Computer Society Press, 2009.
- [73] L. J. LaPadula and D. E. Bell. Secure computer systems: A mathematical model. Technical Report MTR-2547, Volume 2, MITRE Corporation, 1973.
- [74] P. Laud. Handling encryption in an analysis for secure information flow. In *Proc. 12th European Symposium on Programming*, volume 2618 of *Lecture Notes in Computer Science*. Springer-Verlag, 2003.
- [75] P. Lincoln, J. C. Mitchell, M. Mitchell, and A. Scedrov. A probabilistic poly-time framework for protocol analysis. In *Proc. 5th ACM Conference on Computer and Communications Security (CCS'98)*, pages 112–121, 1998.
- [76] A. Lomuscio and B. Woźna. A complete and decidable security-specialised logic and its application to the tesla protocol. In *Proc. 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*, pages 145–152, 2006.
- [77] G. Lowe. An attack on the Needham-Schroeder public-key authentication protocol. *Information Processing Letters*, 56:131–133, 1995.
- [78] G. Lowe. Breaking and fixing the Needham-Schroeder public-key protocol using FDR. In *Proc. 2nd International Workshop on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'96)*, volume 1055, pages 147–166. Springer-Verlag, 1996.
- [79] G. Lowe. Casper: A compiler for the analysis of security protocols. *Journal of Computer Security*, 6:53–84, 1998.
- [80] G. Lowe. Analysing protocols subject to guessing attacks. In *Proc. Workshop on Issues in the Theory of Security (WITS'02)*, 2002.
- [81] H. Mantel. Possibilistic definitions of security — an assembly kit. In *Proc. 13th IEEE Computer Security Foundations Workshop (CSFW'00)*, pages 185–199. IEEE Computer Society Press, 2000.
- [82] D. McCullough. Specifications for multi-level security and a hook-up property. In *Proc. 1987 IEEE Symposium on Security and Privacy*, pages 161–166. IEEE Computer Society Press, 1987.
- [83] D. McCullough. Noninterference and the composability of security properties. In *Proc. 1988 IEEE Symposium on Security and Privacy*, pages 177–186. IEEE Computer Society Press, 1988.
- [84] J. McLean. A general theory of composition for trace sets closed under selective interleaving functions. In *Proc. 1994 IEEE Symposium on Security and Privacy*, pages 79–93, 1994.



- [85] C. Meadows. Formal verification of cryptographic protocols: A survey. In *Advances in Cryptology (ASIACRYPT'94)*, volume 917 of *Lecture Notes in Computer Science*, pages 133–150. Springer-Verlag, 1995.
- [86] C. Meadows. The NRL protocol analyzer: An overview. *Journal of Logic Programming*, 26(2):113–131, 1996.
- [87] M. J. Merritt. *Cryptographic Protocols*. PhD thesis, Georgia Institute of Technology, 1983.
- [88] J.-J. C. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1995.
- [89] D. Micciancio and B. Warinschi. Soundness of formal encryption in the presence of active adversaries. In *Proc. Theory of Cryptography Conference (TCC'04)*, volume 2951 of *Lecture Notes in Computer Science*, pages 133–151. Springer-Verlag, 2004.
- [90] J. Millen and V. Shmatikov. Symbolic protocol analysis with products and Diffie-Hellman exponentiation. In *Proc. 16th IEEE Computer Security Foundations Workshop (CSFW'03)*, pages 47–61. IEEE Computer Society Press, 2003.
- [91] J. Mitchell, A. Ramanathan, A. Scedrov, and V. Teague. A probabilistic polynomial-time calculus for analysis of cryptographic protocols. In *Proc. 17th Annual Conference on the Mathematical Foundations of Programming Semantics*, volume 45 of *Electronic Notes in Theoretical Computer Science*. Elsevier Science Publishers, 2001.
- [92] L. E. Moser. A logic of knowledge and belief for reasoning about computer security. In *Proc. 3rd IEEE Computer Security Foundations Workshop (CSFW'90)*, pages 57–63. IEEE Computer Society Press, 1990.
- [93] Y. Moses. Resource-bounded knowledge. In *Proc. 2nd Conference on Theoretical Aspects of Reasoning about Knowledge (TARK'88)*, pages 261–276. Morgan Kaufmann, 1988.
- [94] R. M. Needham and M. D. Schroeder. Using encryption for authentication in large networks of computers. *Communications of the ACM*, 21(12):993–999, 1978.
- [95] D. M. Nessel. A critique of the Burrows, Abadi and Needham logic. *ACM Operating Systems Review*, 24(2):35–38, 1990.
- [96] C. O'Halloran. A calculus of information flow. In *Proc. European Symposium on Research in Computer Security*, 1990.
- [97] L. C. Paulson. *Isabelle, A Generic Theorem Prover*, volume 828 of *Lecture Notes in Computer Science*. Springer-Verlag, 1994.
- [98] L. C. Paulson. The inductive approach to verifying cryptographic protocols. *Journal of Computer Security*, 6(1/2):85–128, 1998.
- [99] R. Pucella. Deductive algorithmic knowledge. *Journal of Logic and Computation*, 16(2):287–309, 2006.
- [100] F. Raimondi and A. Lomuscio. Verification of multiagent systems via ordered binary decision diagrams. In *Proc. AAMAS'04*, pages 630–637. IEEE Computer Society Press, 2004.
- [101] M. K. Reiter and A. D. Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.
- [102] P. Ryan and S. Schneider. *Modelling and Analysis of Security Protocols*. Addison Wesley, 2000.
- [103] P. Y. A. Ryan and S. A. Schneider. An attack on a recursive authentication protocol: A cautionary tale. *Information Processing Letters*, 65(1):7–10, 1998.

- [104] A. Sabelfeld and A. C. Myers. Language-based information-flow security. *IEEE Journal on Selected Areas in Communications*, 21(1), 2003.
- [105] A. Sabelfeld and D. Sands. Declassification: Dimensions and principles. *Journal of Computer Security*, 17(5):517–548, 2009.
- [106] F. B. Schneider, K. Walsh, and E. G. Sirer. Nexus Authorization Logic (NAL): Design rationale and applications. *ACM Transactions on Information and System Security*, 14(1):1–28, 2011.
- [107] B. Schneier. *Applied Cryptography*. John Wiley & Sons, second edition, 1996.
- [108] G. Smith. On the foundations of quantitative information flow. In *Proc. 12th International Conference on Foundations of Software Science and Computation Structures (FOSSACS’09)*, volume 5504 of *Lecture Notes in Computer Science*, pages 288–302. Springer-Verlag, 2009.
- [109] G. Smith and D. Volpano. Secure information flow in a multi-threaded imperative language. In *Proc. 25th Annual ACM Symposium on Principles of Programming Languages (POPL’98)*, pages 355–364. ACM Press, 1998.
- [110] D. R. Stinson. *Cryptography: Theory and Practice*. CRC Press, 1995.
- [111] S. Stubblebine and R. Wright. An authentication logic supporting synchronization, revocation, and recency. In *Proc. 3rd ACM Conference on Computer and Communications Security (CCS’96)*. ACM Press, 1996.
- [112] D. Sutherland. A model of information. In *Proc. 9th National Computer Security Conference*, pages 175–183, 1986.
- [113] P. Syverson. A logic for the analysis of cryptographic protocols. NRL Report 9305, Naval Research Laboratory, 1990.
- [114] P. Syverson. Towards a strand semantics for authentication logic. In *Proc. 15th Annual Conference on the Mathematical Foundations of Programming Semantics*, volume 20 of *Electronic Notes in Theoretical Computer Science*. Elsevier Science Publishers, 1999.
- [115] P. F. Syverson and S. G. Stubblebine. Group principals and the formalization of anonymity. In *Proc. World Congress on Formal Methods in the Development of Computing Systems*, volume 1708 of *Lecture Notes in Computer Science*, pages 814–833, 1999.
- [116] P. F. Syverson and P. C. van Oorschot. On unifying some cryptographic protocol logics. In *Proc. 1994 IEEE Symposium on Security and Privacy*, pages 14–28. IEEE Computer Society Press, 1994.
- [117] F. J. Thayer, J. C. Herzog, and J. D. Guttman. Strand spaces: Proving security protocols correct. *Journal of Computer Security*, 7(2/3):191–230, 1999.
- [118] B. Toninho and L. Caires. A spatial-epistemic logic and tool for reasoning about security protocols. Technical report, Departamento de Informática, FCT/UNL, 2009.
- [119] M.-J. Toussaint and P. Wolper. Reasoning about cryptographic protocols. In J. Feigenbaum and M. Merritt, editors, *Distributed Computing and Cryptography*, volume 2 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 245–262. American Mathematical Society, 1989.
- [120] R. van der Meyden and N. V. Shilov. Model checking knowledge and time in systems with perfect recall (extended abstract). In *Proc. Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 1738 of *Lecture Notes in Computer Science*, pages 432–445. Springer-Verlag, 1999.

- [121] R. van der Meyden and K. Su. Symbolic model checking the knowledge of the dining cryptographers. In *Proc. 17th IEEE Computer Security Foundations Workshop (CSFW'04)*, pages 280–291. IEEE Computer Society Press, 2004.
- [122] H. P. van Ditmarsch. The Russian Cards problem. *Studia Logica*, 75:31–62, 2003.
- [123] H. P. van Ditmarsch, W. van der Hoek, R. van der Meyden, and J. Ruan. Model checking russian cards. *Electronic Notes in Theoretical Computer Science*, 149(2):105–123, 2006.
- [124] J. van Eijck and S. Orzan. Epistemic verification of anonymity. *Electronic Notes in Theoretical Computer Science*, 168:159–174, 2007.
- [125] P. C. van Oorschot. Extending cryptographic logics of belief to key agreement protocols. In *Proc. 1st ACM Conference on Computer and Communications Security (CCS'93)*, pages 232–243. ACM Press, 1993.
- [126] G. Wedel and V. Kessler. Formal semantics for authentication logics. In *Proc. 4th European Symposium on Research in Computer Security (ESORICS'96)*, volume 1146 of *Lecture Notes in Computer Science*, pages 219–241. Springer-Verlag, 1996.
- [127] J. T. Wittbold and D. Johnson. Information flow in nondeterministic systems. In *Proc. 1990 IEEE Symposium on Security and Privacy*. IEEE Computer Society Press, 1990.